

UNDERSTANDING THE RECOGNITION AND UTILIZATION OF
HOMOGALACTURONAN BY *BACTEROIDES THETA* IOTAOMICRON

Benjamin J. Farnell^{1, 2}
B.Sc., University of Victoria, 2010

A Thesis
Submitted to the School of Graduate Studies
of the University of Lethbridge
in Partial Fulfillment of the
Requirements for the Degree

MASTER OF SCIENCE

Department of Chemistry and Biochemistry
University of Lethbridge
Lethbridge, Alberta, Canada

© Benjamin J. Farnell, 2014

UNDERSTANDING THE RECOGNITION AND UTILIZATION OF
HOMOGALACTURONAN BY *BACTEROIDES THETA IOTA MICRON*

BENJAMIN J. FARNELL

Date of Defense: August 27th, 2014

Dr. D. Wade Abbott Co-supervisor	Adjunct Professor	Ph.D.
-------------------------------------	-------------------	-------

Dr. Steven C. Mosimann Co-supervisor	Associate Professor	Ph.D.
---	---------------------	-------

Dr. Marc R. Roussel Thesis Examination Committee Member	Professor	Ph.D.
--	-----------	-------

Dr. L. Brent Selinger Thesis Examination Committee Member	Professor	Ph.D.
--	-----------	-------

Dr. Mirosław Cygler External Examiner University of Saskatchewan Saskatoon, Saskatchewan	Professor	Ph.D.
---	-----------	-------

Dr. Michael Gerken Chair, Thesis Examination Committee	Professor	Ph.D.
---	-----------	-------

Dedication

To Nana Joy,

You may not have always understood what was happening in the lab or with my research, but that did not stop you from asking about it, or trying to understand. Those are but a few of the conversations I will miss. Thank you for your love and support no matter what the endeavor was.

Abstract

Homogalacturonan (HG) is a structural plant cell wall polysaccharide and a key source of dietary fiber. The human genome does not contain a single enzyme known to be involved in pectin digestion, and therefore, in order to modify HG fibers and potentially extract nutritional value, humans rely on a consortium of symbiotic intestinal bacteria, such as *Bacteroides thetaiotaomicron*, to deconstruct and to ferment this complex carbohydrate into host absorbable products. *B. thetaiotaomicron* contains over 300 predicted carbohydrate active enzymes within its genome that are primarily organized into sugar-selective metabolic pathways called Polysaccharide Utilization Loci PULs (PULs). One such PUL (PUL75: BT4108-BT4124), is activated by HG and is believed to contain enzymes that convert polymerized HG into monosaccharides (GalA). This study reports molecular biology, biochemistry, and functional genomics data that characterize the function of PUL75 gene products involved in HG utilization. Distinct activities for five putative enzymes (BT4108, BT4115, BT4116, BT4119 and BT4123) have been determined on HG substrates. Furthermore, the preferential binding of HG or oligogalacturonides by two defined carbohydrate binding proteins (BT4112 and BT4113) has been established. Based on these findings, a model for the step-wise process of HG recognition, transport and modification by PUL75 is proposed.

Acknowledgements

I would like to thank my supervisor Dr. Wade Abbott for the investment of his time and his lab resources in the completion of this project. I am indebted to him for his guidance, his wisdom and his contagious work ethic. Thank you for being an awesome mentor and a great friend.

I would like to thank my co-supervisor Dr. Steve Mosimann for his insight and his expertise throughout my research project. Thank you for the many discussions on a variety of topics and for teaching me how to replace a computer power source with ease.

To my committee members, Drs. Selinger and Roussel, thank you for your insights and expertise throughout this research project, and particularly during committee meetings. To Dr. Cygler, thank you for acting as my external examiner.

I would like to thank all of the Abbott lab members, past and present, as well as the Mosimann lab members. Carolyn and Kaitlyn, thank you for all the level 2 work done throughout this project. Erin, thank you for the start you gave PUL75 back in the summer of 2012. Justin and Richard, thanks for the company on the late nights and weekends, the good laughs in the office and for enthusiastically listening to all of my ‘newest’ data. Thank you to our collaborators Drs. Harry Gilbert, Sami Tuomivaara, Alisdair Boraston and Benjamin Pluvinaige for their contributions to the research project.

Finally, I would like to thank my wife, Christina and my daughter, Katherine. Your unconditional love, support and patience have been instrumental in the successful completion of this project. In the future, I promise that family outings will not be determined by what experiments can be conducted simultaneously.

Table of Contents

Title Page	i
Signature Page	ii
Dedication	iii
Thesis Abstract	iv
Acknowledgements	v
Table of Contents	vi
List of Tables	viii
List of Figures	ix
List of Abbreviations	xii
Chapter 1: Background and biological relevance	1
1.1 Introduction	1
1.2 Pectic Cell Wall Polysaccharides	6
1.3 Cell Wall Dynamics	9
1.4 Pectinases and the Distal Gut Microbiota	15
1.5 PUL75	24
Chapter 2: Functional characterization of three putative family 1 polysaccharide lyases from PUL75	27
2.1 Introduction	27
2.2 Materials and Methods	33
2.3 Results	40
2.4 Discussion	59
2.5 Conclusion	66

Chapter 3: Elucidating the roles of glycoside hydrolases BT4108 and BT4123 in PUL75 mediated HG deconstruction	67
3.1 Introduction	67
3.2 Materials and Methods	74
3.3 Results	79
3.4 Discussion	93
3.5 Conclusion	100
Chapter 4: Characterization of the PUL75 Sus-like proteins: Insights into substrate binding and affinity	102
4.1 Introduction	101
4.2 Materials and Methods	107
4.3 Results	111
4.4 Discussion	121
4.5 Conclusion	125
Chapter 5: Conclusions and future directions	126
5.1 Overall Summary	126
5.2 Future Directions	130
References Cited	131
Appendix 1	143
Appendix 2	144
Appendix 3	145
Appendix 4	146
Appendix 5	147

List of Tables

Table 1.1 Pectinases within intestinal bacteria that contain multiple genes with predicted or characterized activity on pectic substrates	16
Table 2.1. Matrix digestion template to investigate PL1 complementary activities	37
Table 2.2. Kinetic parameters obtained from BT4115 HG digestions with divalent cations	44
Table 2.3. Mass spectrometry designation of peaks collected from HPAEC-PAD analysis	46
Table 3.1. Matrix digestion template to investigate HG deconstruction by the different enzymes within PUL75	78
Table 3.2. Kinetic parameters obtained for BT4108 enzyme activity on uGalA ₂	81
Table 4.1. PUL75 gene product IDs and localization based on bioinformatic analyses	102
Table 4.2. ITC binding parameters obtained for susE-like proteins	114
Table 4.3. ITC binding parameters obtained for SusD-like proteins	119

List of Figures

Figure 1.1. Pectin structure	4
Figure 1.2. Chemical structure of HG and RG-I	5
Figure 1.3. Basic mechanisms representative of the three classes of polysaccharide active enzymes	12
Figure 1.4. Schematic drawing of PL and GH active sites	14
Figure 1.5. A model of the <i>Bacteroides thetaiotaomicron</i> sus	21
Figure 1.6. CAZyme representatives found in the <i>B. thetaiotaomicron</i> genome	23
Figure 1.7. Visual representation of the seventeen genes that comprise PUL75	25
Figure 2.1. Structure and function of PLs active on HG	29
Figure 2.2. Cloning and recombinant protein purification of the three PL1s from PUL75	41
Figure 2.3. BT4115 pH optimum	42
Figure 2.4. PL1 HG digestion product profiles	43
Figure 2.5. Kinetic analysis of BT4115 with select divalent cations	45
Figure 2.6. HPAEC-PAD analysis of large scale PL1 HG digestions	46
Figure 2.7. Time courses of HG digestion by the PUL75 PL1s: BT4115, BT4116 and BT4119	47
Figure 2.8. Large scale BT4115 HG digestions for uGalA _n generation	48
Figure 2.9. TLC and HPAEC-PAD analysis of PL1 complementary digestions	52
Figure 2.10. HPAEC-PAD and TLC analyses of PL1: BT4115, BT4116 and BT4119 HG time course digestions	53
Figure 2.11. HPAEC-PAD and TLC analyses of PL1: BT4115, BT4116 and BT4119 HG time course digestions	54
Figure 2.12. HPAEC-PAD and TLC analyses of PL1 matrix digestions	55
Figure 2.13. Graphical representation of area under individual unsaturated product peaks from HPAEC-PAD analysis of BT4115, BT4116 and BT4119 HG time course digestions.	56

Figure 2.14. Graphical representation of area under individual unsaturated product peaks from HPAEC-PAD analysis of BT4115, BT4116 and BT4119 HG time course digestions.	57
Figure 2.15. Graphical representation of total area under individual unsaturated product peaks from HPAEC-PAD analysis of BT4115, BT4116 and BT4119 HG time course digestions.	58
Figure 2.16. Cartoon representation of BT4119 predicted domains	65
Figure 3.1. Structure and function of GH28s on pectic carbohydrates	68
Figure 3.2. Structure and function of GH105s on pectic carbohydrates	70
Figure 3.3. Cloning and recombinant protein purification of the two GHs from PUL75	80
Figure 3.4. pH optimum and kinetic analysis of BT4108 (GH105)	81
Figure 3.5. Product profiles of oligogalacturonides produced by GH105 (BT4108) and GH28 (BT4123)	83
Figure 3.6. TLC analysis of GH105 (BT4108) and GH28 (BT4123) complementary time course digestions of unsaturated oligogalacturonides	85
Figure 3.7. TLC analysis of BT4123 (GH28) pH optimum and reducing end assay	87
Figure 3.8. Matrix digestions of HG with BT4108, BT4115 and BT4123 (GH105, PL1 and GH28, respectively)	90
Figure 3.9. pH effect on BT4108, BT4115 and BT4123 matrix digestion of HG	92
Figure 3.10. Proposed subsite orientations for exolytic GH105 (BT4108) and exolytic GH28 (BT4123)	95
Figure 4.1. SusD in complex with maltoheptaose	103
Figure 4.2. SusE in complex with maltoheptaose and maltotetraose	104
Figure 4.3. Cloning and recombinant protein purification of the two SusD-like and susE-like proteins from PUL75	112
Figure 4.4. AGE carbohydrate analysis	113
Figure 4.5. ITC binding isotherm of BT4112 (susE-like) solute binding protein and HG	115

Figure 4.6. ITC binding isotherm of BT4113 (SusD-like) solute binding protein and uGalA ₅ /uGalA ₄ and GalA ₅ /GalA ₄	118
Figure 5.1. Proposed model of PUL75 systematic deconstruction of HG	128
Figure 5.2. TLC analysis of supernatant from wild type <i>B. thetaiotaomicron</i> growth on HG and glucose	129

List of Symbols, Abbreviations and Nomenclature

AceA- aceric acid

Api- apiose

Ara- arabinose

CAZomes- annotated genomes in the CAZy database

CAZy- carbohydrate active enzyme database

CAZymes- carbohydrate active enzymes

CCRC- Complex Carbohydrate Research Center

CE- carbohydrate esterase

DGM- distal gut microbiota

Dha- 2-keto-3-deoxy-D-lyxoheptulosaric acid

DKI-5-keto-4-deoxyuronate

DNSA- 3,5-dinitrosalicylic acid

DP- degree of polymerization

DTT- dithiothreitol

ECF- σ - extra-cellular function σ

EDTA-ethelenediaminetetraacetic acid

Fuc- fucose

Gal- galactose

GalA- galacturonic acid

GalA₂- digalacturonic acid

GalA₃- trigalacturonic acid

GalA_n- saturated oligogalacturonide (where n is the degree of oligomerization)

GH- glycoside hydrolase

GlcA- glucuronic acid

HG- homogalacturonan

HMO- human milk oligosaccharides

HPAEC-PAD- high performance anion exchange chromatography with pulsed amperometric detection

HTCS- hybrid two component system

IMAC- immobilized metal affinity chromatography

IPTG- isopropyl β -D-thiogalactopyranoside

KDO- 2-keto-3-deoxy-D-manno-octulosonic acid

LPS- lipopolysaccharide

MW- molecular weight

MWM- molecular weight marker

NR- non-reducing end

NTA- nickel (II) nitriloacetic acid

OD- optical density

PCR- polymerase chain reaction

PDB- protein data bank

PEG- polyethelene glycol

PEMI- pectin methylesterase inhibitor

PIMS- pectin integrity monitoring system

PL- polysaccharide lyase

PME- pectin methylesterase

PUL- polysaccharide utilization locus

R- reducing end

RG-I- rhamnogalacturonan I

RG-II- rhamnogalacturonan II

Rha- rhamnose

RT- room temperature

SDS-PAGE- sodium dodecyl sulfate polyacrylamide gel electrophoresis

SEC- size exclusion chromatography

sus- starch utilization system

Sx- SephadexTM

TLC- thin layer chromatography

TM- transmembrane

TPRs- tetratrico peptide repeats

Tris- 2-Amino-2-hydroxymethyl-propane-1,3-diol

uGalA- unsaturated galacturonic acid

uGalA_n- unsaturated oligogalacturonide (where n is the degree of oligomerization)

u- unsaturated

WGS- whole genome sequencing

Xyl- xylose

Chapter 1-

Background and biological relevance

1.1 Introduction

The study of carbohydrates in biological systems (glycobiology) is an expanding area of research and discovery. Prior to the 1960's, carbohydrates were considered exclusively to have a passive role as either energy sources (e.g. glucose and starch) or structural components (e.g. cellulose). More recently carbohydrates have been implicated in more dynamic roles such as signaling molecules¹ and influence of distal gut microbiota (DGM) composition². Unlike DNA, carbohydrates do not replicate, nor do they catalyze complex reactions like enzymes; however, their structural diversity is key to their roles within biological systems³. The structural diversity of carbohydrates is in stark contrast to that of DNA and proteins. Namely, DNA and proteins are both linear molecules with one type of linkage. In carbohydrates, individual monosaccharides can be linked in a linear or branched pattern, with some carbohydrates (eg. starch) employing both types of linkages. Furthermore, these linkages (glycosidic bonds) are classified as either α - or β -linkages, based on the anomeric carbon configuration. The ability to chemically modify resident hydroxyl groups of a monosaccharide further increases both diversity and structural complexity within the carbohydrate family.

The plant cell wall serves as a key example of carbohydrate diversity and complexity. Unlike most animals which possess a skeletal structural support system, plants rely on a complex biomolecular network called the plant cell wall. This network is required to be rigid enough to support immense weight, yet pliable to allow for growth and expansion in

various stages of development. The plant cell wall is divided into two main components referred to as the 'primary' and 'secondary' wall⁴. Both walls are predominantly composed of the polysaccharides cellulose, hemicellulose (matrix polysaccharides) and pectin⁵. Cellulose is a pure, unbranched and unmodified β -1-4-glucan polymer. These polymers interact to form higher-order structures called microfibrils and fibrils that are primarily crystalline due to dense inter- and intramolecular hydrogen bonding, and the exclusion of water⁴. Cellulose is synthesized at the plasma membrane from nucleotide sugar substrates and is a component of both the primary and secondary wall⁶.

Hemicelluloses differ in composition from cellulose but are structurally analogous in that they are connected through β -linkages. The main classes of hemicelluloses include xylans, mannans and glucans, which are homopolymers of xylose, mannose, and glucose, respectively. Variations in structure do exist. Glucomannans for example, have a backbone of randomly dispersed β -1-4-linked glucose and mannose⁵. These different backbones can be extensively decorated with a variety of sugars and acetyl groups which account for the non-crystalline nature of these polymers. Much like cellulose, hemicelluloses are present in both the primary and secondary wall; however, they are synthesized in the Golgi from the corresponding nucleotide sugar substrates⁶.

Pectin is a plant cell wall structural polysaccharide within the primary cell wall and the middle lamella, which punctuates the junctions between primary walls of neighboring cells, and participates in intercellular connections⁷. In 1790, Nicolas Vauquelin first identified pectin in apples; however, it wasn't until 1825 that Henri Braconnot named it after the Greek word *pēktikós*, which means to congeal or solidify. In addition to being

found in the cell walls of plants and fruits, it is also found in the cell walls of some freshwater and marine algae^{8, 9}. Pectin is the most complex carbohydrate found in nature due to stereochemical diversity of glycosidic bonds that link a variety of common and rare carbohydrate subunits. A defining feature of all pectins is that it displays a high D-galacturonic acid (GalA) content¹⁰ (Figure 1.1). GalA adopts a ⁴C₁ conformation and is structurally analogous to D-galactose (Gal) with an equatorial C6 that has been oxidized into a uronic acid (Figure 1.2A-C). Each GalA moiety (pKa 3.5) therefore contains an inherent negative charge at physiological pH.

As early as the 1750's, pectin was being used as a gelling agent in jams and jellies made from apples and currants, as described in the 'London Housewife's Companion'. The food processing industry has continued this most common use of pectin as a gelling agent in jams, jellies and preserves¹¹. More recently with the trend in low-calorie foods, pectin has been used as a fat and sugar replacement, and it has even been used as an additive to retard crystal formation in frozen foods. On the medicinal side, pectin has been used to regulate cholesterol levels in the blood as well as a chelating agent for toxic metals¹¹. More recently, the effects of dietary fiber, of which pectin is a main component, on the composition of the distal gut microbiota have been investigated².

Within the plant cell wall, pectin is divided into three classes of distinct pectic polysaccharides: homogalacturonan (HG), rhamnogalacturonan-I (RG-I) and rhamnogalacturonan-II (RG-II)¹² that vary in size, branching, and function. HG, RG-I, and RG-II are believed to be found as an interconnected macromolecule within the plant

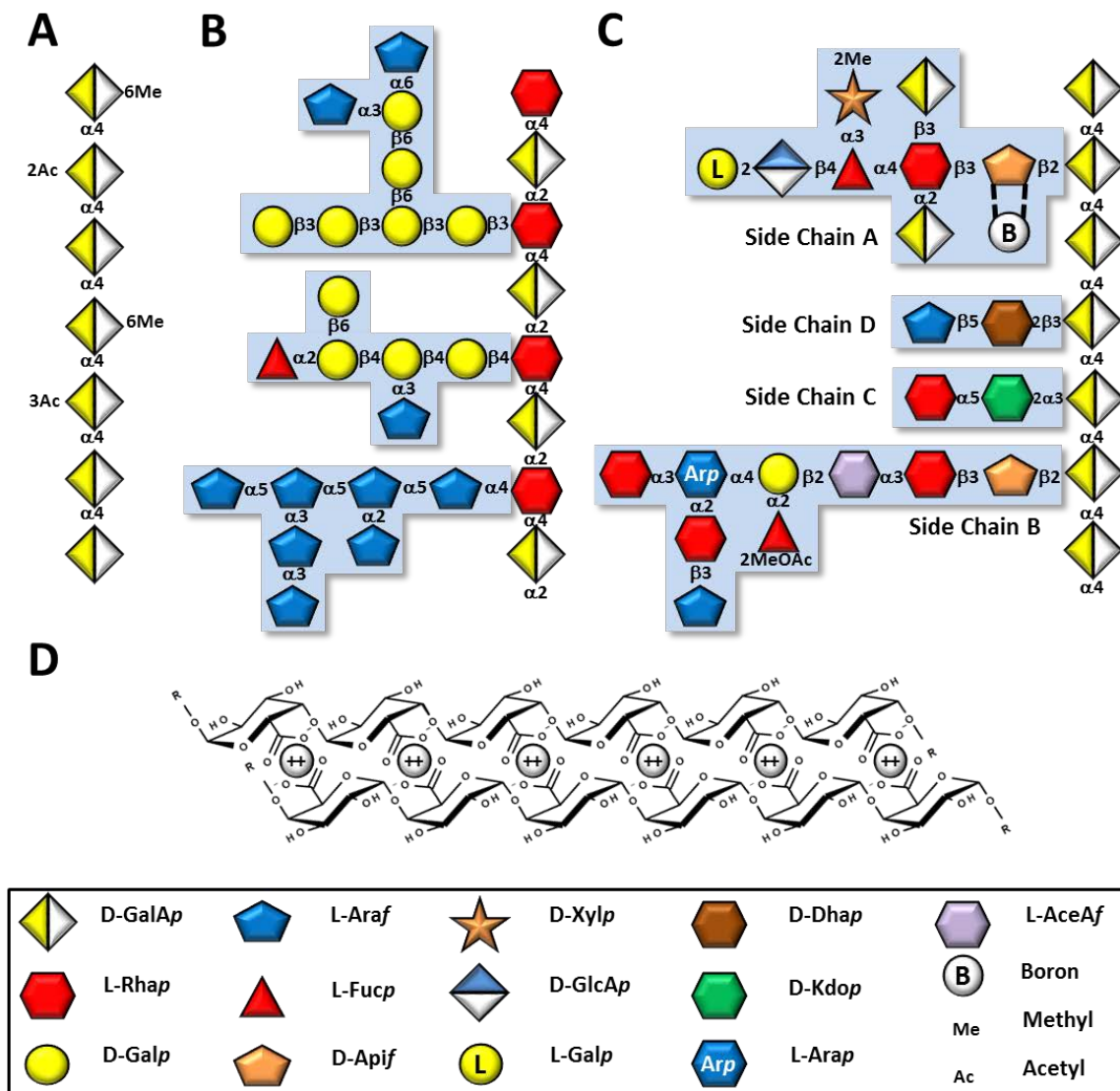


Figure 1.1: Pectin structure. Schematic representations of the three main pectic polysaccharides (A) Homogalacturonan, (B) Rhamnogalacturonan-I and (C) Rhamnogalacturonan-II. Sugars are represented using the standard symbol nomenclature. Isomerization is indicated by D and L followed by *p* (pyranose) or *f* (furanose) to indicate ring configuration. Sugar linkages are indicated between linked residues. Carbohydrate nomenclature is as follows: GalA: Galacturonic acid; Rha: Rhamnose; Gal: Galactose; Ara: Arabinose; Fuc: Fucose; Api: Apiose; Xyl: Xylose; GlcA: Glucuronic acid; Dha: 2-keto-3-deoxy-D-lyxoheptulosaric acid; KDO: 2-keto-3-deoxy-D-manno-octulosonic acid; and AceA: Aceric acid. (D) Egg-box model. The intermolecular coordination of a calcium divalent cation (++), by the C5 uronic acid group of proximal GalA moieties creates a tightly packed structure, wherein the calcium represents the egg, 'boxed' in between to HG polymers.

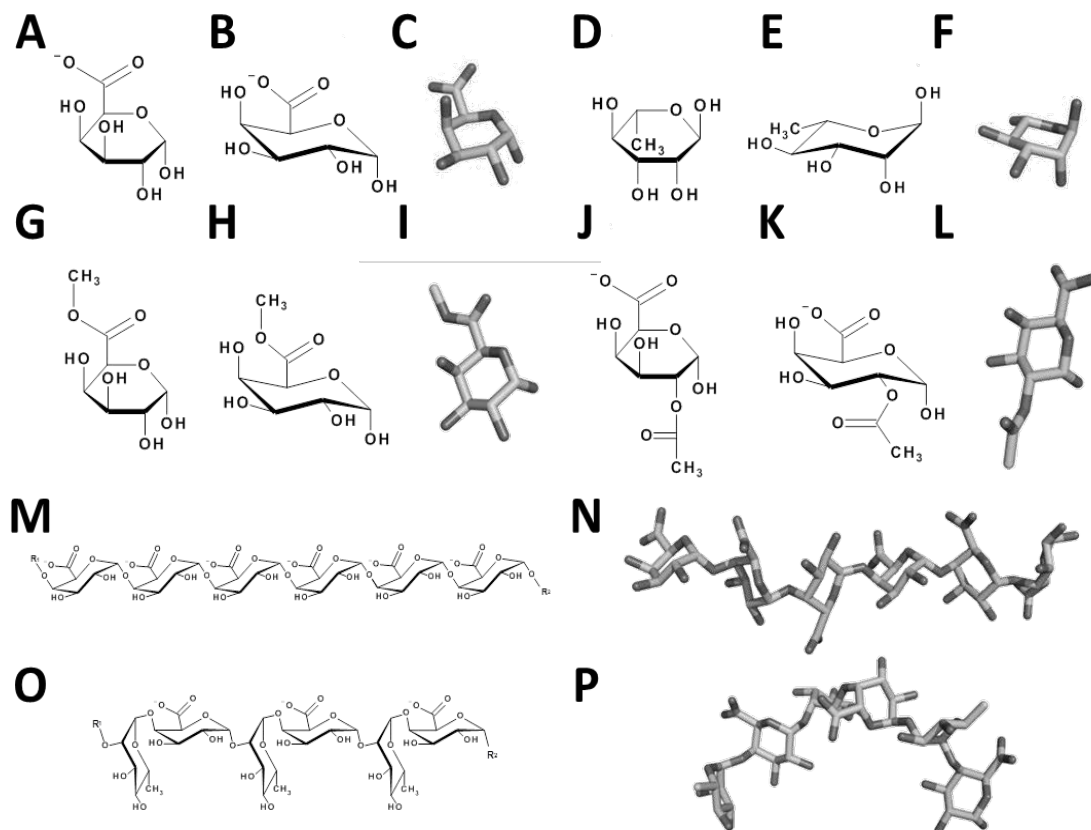


Figure 1.2: Chemical structure of HG and RG-I. α -D-GalA displayed as a Haworth (A) and chair (B) projection. (C) Three-dimensional structure of α -D-GalA extracted from the PL1 structure from *E. chrysanthemi* EC16 (PDB ID: 2EWE)¹³. α -L-Rha displayed as a Haworth (D) and chair (E) projection. (F) Three-dimensional structure of α -L-Rha extracted from the PL4 structure from *A. aculeatus* KSM 510EC16 (PDB ID: 3NJV)¹⁴. Methyl esterified α -D-GalA displayed as a Haworth (G) and chair (H) projection. (I) Three-dimensional structure of methyl esterified α -D-GalA extracted from the CE8 structure from *D. dadantii* 3937 (PDB ID: 2NST)¹⁵. Acetyl esterified α -D-GalA displayed as a Haworth (J) and chair (K) projection. (L) Three-dimensional model of acetyl esterified α -D-GalA built from an β -N-acetylgalactosamine scaffold and validated for bond angles and distances using COOT¹⁶. (M) Schematic representation of HG. (N) Three-dimensional structure of a HG hexasaccharide extracted from the PL1 structure of *E. chrysanthemi* EC16 (PDB ID: 2EWE). (O) Schematic representation of RG-I. (P) Three-dimensional structure of a RG-I hexasaccharide extracted from the PL4 structure of *A. aculeatus* KSM 510EC16 (PDB ID: 3NJV).

cell wall (Figure 1.1A-C)¹⁷. There appears to be multiple levels of covalent crosslinking that contribute to this network, which include but are not limited to, backbone glycosidic linkages, calcium crosslinking¹⁸, borate ester coordination¹⁹ and covalent linkages to phenols (lignin), proteins, and possibly other compounds yet to be discovered²⁰. Elucidating the specific assemblies and the degrees of polymerization of each pectic domain remains a difficult task as the ‘native’ pectin structure is disrupted by chemical and enzymatic treatments required to extract it from the primary cell wall. Despite these limitations, distinct and integral functions have been correlated with pectic networks, including, cell-cell adhesion via HG cross-linking and RG-II dimerization^{20, 21}, chemical signaling²⁰, growth and development, fruit development (ripening)^{22, 23} and plant defense responses²⁴⁻²⁶.

1.2 Pectic Plant Cell Wall Polysaccharides

1.2.1 Homogalacturonan— HG is synthesized by α -1,4-galacturonosyltransferases, such as GAUT1²⁷, which create highly polymerized fibers of α -1,4-linked GalA (also referred to as polygalacturonic acid and pectate)²⁴. The glycosidic bonds of HG are connected through a C1 axial - C4 axial ‘accordion-like’ structure that is uncommon in other polysaccharides (Figure 1.2M-N). This accounts for the ‘twisted’ linear shape of the HG polysaccharide (Figure 1.2N), as opposed to a more helical structure found in other carbohydrate polymers²⁸ (Figure 1.2P). HG is the backbone of pectin, accounting for greater than 60% of the total pectin assembly^{24, 29}. This backbone can be chemically modified by either methyl esterification (C6, Figure 1.2G-I) and/or acetylation (O2 or O3, Figure 1.2J-L)²⁴. The pattern of methyl esterification and/or acetylation varies

between plant species, suggesting that the degree of chemical modification is related to developmental and tissue-specific phytophysiology³⁰. Contiguous regions of HG (>10 residues) lacking these chemical modifications are capable of forming Ca^{2+} salt bridges between the negative charges of uronate groups, stabilizing a defined higher order structure, referred to as the ‘egg-box model’ (Figure 1.1D)^{18, 20}. ¹³C NMR experiments have shown that the gelatinous HG (egg-box model) adopts a 2₁ helical conformation (two residues per turn), whereas dried HG adopts a 3₁ helical conformation³¹. This model contributes to dense packing of HG into pectic gels, with ~70% of the pectate accounted for in this gel form. This gelling property also accounts for the extensive use of pectin in the food processing industry.

1.2.2 Rhamnogalacturonan-I— RG-I is unique amongst the pectic polysaccharides as its backbone is comprised of a repeating disaccharide of GalA and Rhamnose (Rha) [4)- α -D-GalA-(1,2)- α -L-Rha-(1,)]_n^{29, 32, 33} (Figure 1.2D-F, O-P). Rha is a stereoisomer of 6-deoxy-L-mannose that displays a C2 axial hydroxyl in the ¹C₄ conformation. The axial-axial linkage of GalA (O4) to Rha (O2) results in the three-dimensional structure of RG-I adopting a curved helix (Figure 1.2O-P), which is strikingly different than the twisting linear structure of HG (Figure 1.2M-N). Similar to HG, the backbone GalA residues may be hyperacetylated at the O2 and O3 positions and 25-80% of the Rha residues are decorated by branching at the O4 position¹². These decorations include linear or branched patterns of defined polysaccharides: α -1,5 (arabinans), and β -1,3 or β -1,4 (galactans, Figure 1.1B)^{24, 34, 35}. RG-I side-chains can contain further branching at certain positions (O2 and O3 for arabinans and O3 and O6 for galactans) by the following

sugars; arabinose, arabinan, galactan, arabinogalactan³⁴. Secondary substitutions of the linear or branched polysaccharides from the main GalA-Rha backbone increase its complexity and lead to a wide diversity of possible RG-I structures. Due to these extensive decorations, RG-I is commonly referred to as the ‘hairy-region’ of pectin. Interestingly, the structure of RG-I is not strictly conserved between tissues and species, but rather its side-chains appear to be developmentally and differentially regulated^{22, 36, 37}.

1.2.3 Rhamnogalacturonan-II— RG-II, which accounts for ~10% of pectin²⁹, is comprised of an HG backbone (~7-9 α -1,4-linked GalA) with four well-defined side-chains²⁰ (labeled A-D) (Figure 1.1C). Side-chains A (an octasaccharide) and B (a nonasaccharide) are linked to the HG backbone at the O2 position. Side-chains C and D are both disaccharides and are linked to the HG backbone at the O3 position. These four defined and well conserved side chains add to the complexity of RG-II molecule by presenting 12 different types of monosaccharides, including the following rare sugars: 2-O-methylxylose, 2-O-methylfucose³⁸, aceric acid³⁹, 2-keto-3-deoxy-D-lyxoheptulosaric acid (Dha)⁴⁰, and 2-keto-3-deoxy-D-manno-octulosonic acid (KDO) (Figure 1.1C)⁴¹.

In addition to the abundance of different carbohydrate subunits, RG-II also displays a total of 21 different linkages between these 12 identified subunits. Despite this structural diversity (i.e. sugars and linkages), RG-II is highly conserved between plant species^{12, 19}. RG-II generally exists as an RG-II dimer that is cross-linked by a bidentate borate diester between apiose residues in side-chain A²¹, which covalently crosslinks two distinct RG-II molecules and fortifies the pectin network²¹. These conserved features of RG-II structure

play a critical function in plant growth and development, as minor modifications of the RG-II structure have shown near fatal effects on plant growth⁴².

1.3 Cell Wall Dynamics

Though a common characteristic of plants, cell wall structure varies by source and can be altered over time. Lignification is the hardening process that occurs in the secondary wall which renders the wall resistant to compressive forces, while restricting passage of small molecules⁴. In some cases a softening of the cell wall or ripening, occurs instead. Ripening (with other associated chemical changes) occurs due to the regulated modification of polysaccharides found within the primary cell wall and middle lamella. The majority of these polysaccharide modifications are the result of secreted carbohydrate active enzymes (CAZymes)^{23, 43}. Initially, these enzymes target the HG rich middle lamella, which results in a loss of intercellular connections⁷. This is followed by enzymatic modifications to the cell wall, resulting in a weakened cell wall structure and network⁴⁴. These modifications allow for the hard and acidic unripe tissue to transform into an attractive, sweet, fragrant and soft fruit²³. Ripening is a limiting factor in the distribution of fruit worldwide, and thus an area of financial significance to the fruit industry.

Dynamic modification of the plant cell wall during infectious disease is catalyzed by CAZymes produced by various organisms. Bacterial, fungal, and insect pathogens are known to contain enzymes, such as pectin methylesterases (PMEs), that degrade the plant cell wall, a process that can lead to disease and even plant death (soft-rot)⁴⁵. In order to defend against invading pathogens, plants deploy a coordinated immune cascade. A

primary line of defense involves an immune protein, referred to as pectin methylesterase inhibitor (PMEI), which disrupts the function of PMEs^{46, 47}. PMEIs, bind to the active site of pectin methylesterases (PME) secreted by phytopathogens, which renders the enzyme inactive⁴⁸. PMEs are a primary virulence factor during infection and function upstream of depolymerases such as pectate lyases and polygalacturonases. If the pectic network becomes compromised, plants have an innate immune response that is elicited by the release of oligogalacturonides, which includes accumulation of reactive oxygen species⁴⁹ and pathogenesis-related proteins^{50, 51}. Recently the existence of a pectin integrity monitoring system (PIMS) has been proposed²⁶ which are regions of HG that, when cleaved by invading species, act as a signal to activate plant innate immunity.

1.3.1 Enzymatic modification of pectic polysaccharides— Despite its abundance in nature, there are relatively few enzyme families known to cleave the glycosidic linkages within the pectic backbones of HG and RG-I. Ultimately, the cause of this observation is not known, however it may lie with the inherent properties of the substrate itself (e.g. stability, steric constraints, and charge potential) or be a result of the selective pressures of convergent evolution of pectinases. Interestingly, when present in the genomes of pectinolytic bacteria, pectinases are often found in multiple copies. This observation suggests two plausible explanations; individual genes are differentially regulated or display a preferential activity, which has been ‘tuned’ towards substrates that may vary in their degree of esterification, polymerization or carbohydrate composition (Figure 1.2)^{52, 53}. Currently there are two distinct and mechanistically different enzyme classes known to cleave the glycosidic linkages within pectic sugars; polysaccharide lyases (PLs) and glycoside hydrolases (GHs).

1.3.2 Polysaccharide lyases— Polysaccharide lyases (PLs) are a group of enzymes that employ a β -elimination reaction to cleave their uronic acid-containing polysaccharide substrate⁵⁴. PL activity results in the generation of both a hexenuronic acid moiety (non-reducing end) and a new reducing end⁵⁴(Figure 1.3A). The β -elimination mechanism can be broadly described by these three main events: (i) proton abstraction from the C5 sugar residue by a basic amino acid, (ii) stabilization of the resulting anion and (iii) cleavage of the C4:O4 scissile bond. Substrate coordination and recognition is often facilitated and dependent on a divalent cation or positively charged amino acid in proximity to the active site. There is an understanding that these cations/residues also play a role in the catalytic mechanism, however their exact roles remains poorly understood. For a more exhaustive description of specific PL families active on the pectic backbone see Chapter 2.

1.3.3 Glycoside hydrolases- Glycoside hydrolases (GHs) are the largest group of CAZymes identified to date⁴³. This group is composed of enzymes that utilize water to facilitate the cleavage of glycosidic linkages between sugars as well as O- and N- linked glycans⁵⁵. GH activity results in the generation of a hemiacetal or hemiketal and an aglycone (non-sugar moiety/leaving group). Consistent with PLs, GHs also generate a new reducing end after hydrolysis (Figure 1.3B).

GHs have two broadly defining reaction mechanisms; inverting and retaining. The nomenclature of these two mechanisms refers to the stereochemistry at the anomeric carbon after cleavage. The inverting mechanism occurs via a one step, single-displacement mechanism whereby two amino acid side chains (typically glutamic or aspartic acid) assist in acting as a general acid and general base.

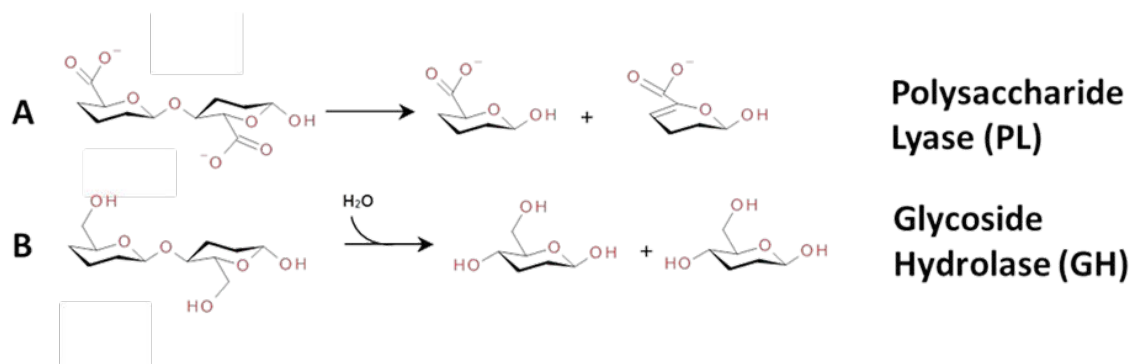


Figure 1.3: Basic mechanisms representative of two classes of polysaccharide active enzymes

The retaining mechanism, on the other hand, occurs via a two-step, double-displacement mechanism that involves a glycosyl-enzyme intermediate. Two amino acid side chains (typically glutamic or aspartic acid) assist in acting as an acid/base or a nucleophile. For a more exhaustive description of specific GH families active on the pectic backbone see Chapter 3.

For both classes of enzyme, a nomenclature to describe the active site in schematic representations has been proposed⁵⁶. Within the active site, there are a minimum number of subsites that must be populated in order to obtain activity. Subsites are divided into negative and positive (left- and right-hand side, respectively) with the glycosidic bond to be cleaved (scissile bond) consistently positioned between the -1 and +1 subsites. By convention, the substrate is drawn with the non-reducing end on the left, and reducing end on the right. A schematic representation of the active site for both a PL and GH, and the resulting products are presented in Figure 1.4.

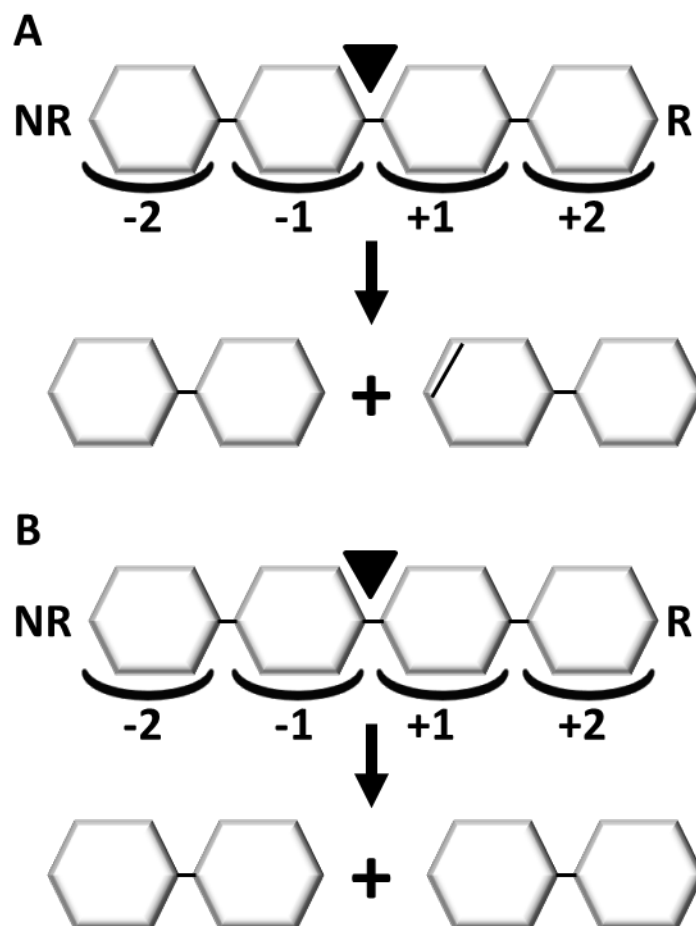


Figure 1.4: Schematic drawing of GH and PL active sites. (A) GH enzyme active site and subsequent products after cleavage of the scissile bond. (B) PL enzyme active site and subsequent products after cleavage of the scissile bond. By convention, the non-reducing end (NR) of the substrate is on the left, with the reducing end (R) on the right. The scissile bond is indicated by the black arrow and is always positioned between the -1 and +1 subsites. The number of subsites and their configuration are enzyme dependent.

1.4 Pectinases and the Distal Gut Microbiota (DGM)

The majority of our understanding and characterization of pectinases has come from a number of model organisms that display activity toward pectic polysaccharides^{57, 58}. These model organisms from the Enterobacteriaceae (i.e. *Erwinia spp.*, *Dickeya spp.*, and *Pectobacterium spp.*)^{53, 59} and pectinolytic fungi (e.g. *Fusarium graminearum*, *Aspergillus sp.*)^{60, 61} represent a number of the ‘soft-rot’ pathogens that cause devastation to commercial agricultural crops worldwide. The genomes of these pectinophiles contain an arsenal of pectinolytic enzymes well adapted to deconstruct the intact plant cell wall. These pectinolytic enzymes vary in regulation, cellular target and activity. Pectinolytic intestinal bacteria, on the other hand, have more limited and specialized pathways for pectin utilization. These pectinolytic intestinal bacteria represent an emerging field for the further understanding and characterization of pectinases. Among the annotated genomes in the CAZy database (i.e. CAZomes), the *Bacteroides* genus displays a large number of genes belonging to pectinase families (Table 1.1). Interestingly, there are a number of pectinolytic strains of bacteria, that when compared to members of their own species, display augmented levels of predicted pectinase genes. This observation highlights the diversity and metabolic potential of select pectinolytic intestinal bacteria. Furthermore, it reinforces the need for next generation metagenomic initiatives, such as the Human Microbiome Project⁶², that provide the sequencing depth to define these relationships.

Table 1.1: Pectinases within intestinal bacteria that contain multiple genes with predicted or characterized activity on pectic substrates⁴³

SPECIES	PL1	PL2	PL3	PL4	PL9	PL10	PL11	PL22	GH28	GH105	CE8	CE12	Total
BACTEROIDES													
<i>B. salanitronis</i> DSM 18170 ¹	3					3	2		11	6	5	4	34
<i>B. thetaiotaomicron</i> VPI-5482	5				2	1	1		9	7	3	4	32
<i>B. vulgatus</i> ATCC 8482	2					2	3		13	7	4	5	36
<i>B. xylanisolvens</i> XB1A	5				1	1	4		9	5	2	2	29
ENTEROBACTER													
<i>E. aerogenes</i> EA1509E								1	1	2	1		5
<i>E. aerogenes</i> KCTC 2190									1	2	1		4
<i>Enterobacter</i> sp. 638		1			1			1	1	1	1		6
ENTEROCOCCUS													
<i>E. casseliflavus</i> EC20									1	2	1	1	5
<i>E. faecium</i> Aus0004					1				2	2	1	1	7
<i>E. mundtii</i> QU 25 QU25									1	3	1	1	6
KLEBSIELLA													
<i>K. oxytoca</i> E718		1			1			2	3	2	2		11
<i>K. pneumoniae</i> 342 ²								1	1	2	1		5
ROSEBURIUM													
<i>R. hominis</i> A2-183									2	1	1		4
<i>R. intestinalis</i> M50/1									1	1	1		3
YERSINIA													
<i>Y. enterocolitica</i> 8081 ²		2						1	1				4
<i>Y. pseudotuberculosis</i> YPIII ²		2						1		1	1		5
OUTLIERS													
<i>Eubacter. eligens</i> ATCC27750	3				4				3	2	2	2	16
<i>Faecalibacter. prausnitzii</i> SL3/3									3	1	1		5
<i>Salmonella bongori</i> N268-08 ²									1	1	1		3

¹*B. salanitronis* DSM 18170 was isolated from poultry.

²Human pectinolytic pathogens

The human genome does not contain the required enzymes to digest pectic polysaccharides (HG, RG-I and RG-II). In order to unlock the caloric potential contained in the glycosidic linkages of these polysaccharides we rely on a consortium of symbiotic bacteria⁶³. The anaerobic fermentation of these enzymatically cleaved polysaccharides by the DGM produces a number of host-absorbable byproducts (e.g. acetate, propionate, and butyrate)⁶⁴. Compared with the intact cell wall that is targeted by saprobes and phytopathogens, the pectic substrates metabolized by intestinal bacteria have been ‘pretreated’ by the upper stages of digestion. Mastication and acid hydrolysis of the pectic substrates present the intestinal bacteria with a more accessible substrate and may provide a rationale for the differences observed in pectin utilization pathways between saprobes and phytopathogens on the one hand, and intestinal bacteria on the other.

1.4.1 Bacteroides spp.— *Bacteroides* is a genus of ubiquitous glycophilic bacteria found in terrestrial, marine, and host-intestinal habitats. In adult humans, members of the genus *Bacteroides* account for ~25% of the total bacterial population in the intestines⁶⁵. The *Bacteroides* belong to the family *Bacteroidaceae*, order *Bacteroidales*, class *Bacteroides*, phylum *Bacteroidetes*, and *Bacteroidetes/Chlorobi* group (Superphylum) (www.ncbi.nlm.nih.gov/Taxonomy). Phylogenetic analyses indicate that a large distance exists between the Bacteroidetes and Protobacterium or Firmicutes phylum, indicating that the Bacteroidetes diverged early in the evolution of Bacteria.

Bacteroides spp. are anaerobic, non-spore forming, bile-resistant, gram negative rod shaped bacteria. The lipopolysaccharide (LPS) layer of *Bacteroides* is unique from other bacteria in both structure (inclusion of sphingolipids and a phosphorylated KDO) and in

the type of host immune response initiated. Typically bacterial LPS incite an endotoxic response. *Bacteroides* LPS incites a non-LPS response, and is 10-1000 fold less toxic than *E. coli* LPS⁶⁶. The LPS can be both smooth or rough, depending on the length and branching patterns present. Within *Bacteroides* there are reports of the presence of a complete capsule, a partial capsule or no capsule⁶⁷. The *Bacteroides* capsule incites a unique host response, abscess formation, which can lead to further complications and result in death⁶⁸. Within adult human intestines⁶⁹, most *Bacteroides* spp. maintain a complex and generally beneficial (mutualistic) relationship with the host, however, *Bacteroides fragilis* is regarded as the most virulent strain among *Bacteroides* spp., representing the most frequent isolate from clinical specimens⁶⁸.

In 2003, *Bacteroides thetaiotaomicron* type strain VPI-5482 isolated from the feces of a healthy human⁷⁰ became the first sequenced member of the *Bacteroides* family. This was quickly followed by the *B. fragilis* NCTC 9343 genome from 2004-2005^{71, 72}. This strain was isolated from two different patients, both of whom were suffering from abdominal infections. The completion of these WGS projects and the subsequent proteomic analyses provided a wealth of information to better understand these bacteria and their respective roles within the human gut as both mutualistic (*B. thetaiotaomicron*) and pathogenic (*B. fragilis*) organisms. Two main highlights from these projects were the identification of extensive DNA inversions that control and regulate a number of gene clusters and the identification of gene clusters responsible for sensing and utilizing nutrients. These DNA inversions were found in a number of polysaccharide biosynthesis regions of the genome facilitating the ability to modulate surface polysaccharides, which has been demonstrated as a mechanism utilized by the bacteria to evade host immune

response⁷³. Similarly, the presence of gene clusters responsible for sensing and utilizing nutrients (host and plant derived glycans), may provide a further advantage to the *Bacteroides* spp. within the dynamic and competitive distal gut environment.

Metabolic pathways in *Bacteroides* spp. are organized into dedicated gene clusters, referred to as a Polysaccharide Utilization Locus (PUL). PULs are a conserved genomic feature of *Bacteroides* spp. and represent dedicated and independently regulated carbohydrate utilization pathways⁷⁴. Though PULs are ‘self-contained’ and respond to discrete carbohydrate signals, there are certain requirements necessary for their proper function: regulation and sensing, transport and energy harvest.

1.4.2 Regulation and sensing— Regulation and sensing are most commonly the role of either the extra-cellular function σ (ECF- σ)/anti- σ pairs or the Hybrid Two Component System (HTCS)⁷⁵. HTCS combine the two domains of the classical two-component environmental sensors (a sensor kinase and a response regulator) into a single polypeptide. This single polypeptide has five predicted domains that span the cellular membrane, an N-terminal periplasmic sensor and four conserved cytoplasmic domains; a histidine kinase, an ATPase, a regulator, and a DNA-binding domain⁷⁶. In the periplasmic space, binding of the inducer to the sensor translates gene expression and activation of the discrete PUL.

1.4.3 Carbohydrate transport— Transport is carried out by the ‘Sus-like’ proteins (SusCDEF) (Figure 1.5). These proteins are homologs of starch binding proteins found in the starch utilization system (sus)^{77, 78}. SusC-like proteins are alleged to be TonB-dependent, β -barrel-type outer membrane proteins. As such, they are thought to

participate in the energy-dependent transport of the bound polysaccharide into the periplasmic space⁷⁹. SusD-like proteins are predicted to be outer membrane lipoproteins due to the presence of an N-terminal lipid⁸⁰. They bind directly to specific glycans and aide in the capture and delivery of glycans to the SusC-like transporter^{81, 82}. SusE- and SusF-like proteins are ancillary binding proteins that aid in the initial recruitment of target polysaccharide to the cell surface, often representing the initial steps in the metabolic process⁸³.

1.4.4 Energy harvest— Deconstruction of the target polysaccharide is accomplished by the co-operation and co-ordination of two main classes of enzymes: Glycoside Hydrolases (GH) and Polysaccharide Lyases (PL), with the aid of a third class Carbohydrate Esterases (CE), as required. The specific activities of the enzymes present in a PUL are responsible for cleavage of targeted glycosidic linkages, resulting in the deconstruction of the target polysaccharide by the complementary enzymes. (Figure 1.5)

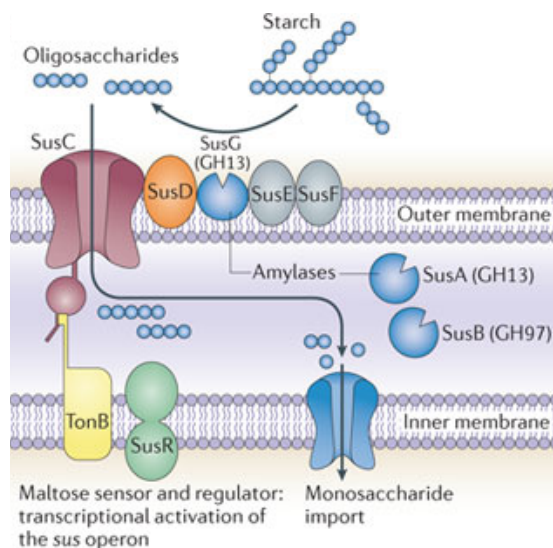


Figure 1.5⁶³ A model of the *Bacteroides thetaiotaomicron* Sus. Starch binding is initiated by SusD, SusE and SusF, initial deconstruction is carried out by SusG, and oligosaccharides are transported into the periplasm via SusC in concert with the inner-membrane protein TonB. In the periplasm, malto-oligosaccharides are further deconstructed to glucose by another GH13 enzyme, SusA, and a GH97 enzyme, SusB. The presence of liberated maltose is sensed in the periplasm by the inner-membrane-spanning regulator SusR, which activates expression of the other Sus proteins.

1.4.5 Human gut colonization— The colonization of *Bacteroides* spp. in the human gut may occur as early as birth, as the newborn passages through the vaginal canal, inheriting their mother's microbiota⁸⁴. From this initial exposure *B. thetaiotaomicron* may have a competitive advantage in the early colonization of the distal infant gut as it has been shown to utilize human milk oligosaccharides (HMOs) via induction of a mucus utilization pathway⁸⁵. As the infant is weaned, progressing to a more complex diet rich in plant glycans, *B. thetaiotaomicron* is primed for this dietary shift, due to the large presence of PULs within its genome, providing a distinct advantage over other resident bacteria.

Within the DGM, *B. thetaiotaomicron* has become the model symbiont in the study of host interactions. A number of inherent factors has led to this; genetically

manipulatable⁸⁶, easy to culture and a predominant member of the distal intestinal microbiota of both mice and humans⁶⁹. Furthermore, the extensive array of CAZymes (Figure 1.6) provides a wealth of unique carbohydrate food sources for the bacterium. As such, continued effort to further describe *B. thetaiotaomicron*'s symbiotic role within the DGM remains a high priority.

Glycoside Hydrolase Family		2	3	13	16	18	20	23	25	27	28	29	30	31	32	33	35	36	38	42	
Number of sequences		32	10	8	4	12	14	3	1	5	9	9	3	6	4	2	3	4	2	1	
		43	51	53	57	66	67	73	76	77	78	84	88	89	92	93	95	97	99	105	
		33	4	1	1	1	1	1	10	1	6	1	4	3	23	1	5	10	1	7	
		106	109	110	115	123	125	127	130	NC											
		3	2	2	2	1	3	3	4	2											
GlycosylTransferase Family		2	3	4	5	8	9	14	19	25	28	30	32	35	51	NC					
Number of sequences		38	1	23	1	1	1	3	1	1	1	1	3	2	4	6					
Polysaccharide Lyase Family		1	8	9	10	11	12	13	15	NC											
Number of sequences		5	3	2	1	1	2	1	1	1											
Carbohydrate Esterase Family		1	2	4	6	7	8	9	11	12											
Number of sequences		2	1	3	1	1	2	3	1	4											
Carbohydrate-Binding Module Family		6	20	32	35	50	51	57	58	62	NC										
Number of sequences		1	2	13	2	2	1	1	1	1	2										

Figure 1.6: CAZy representatives found in the *B. thetaiotaomicron* VPI-5482 genome. The five main CAZy groups are indicated on the left, with the family number represented in the top box, and number of predicted family members found in the *B. thetaiotaomicron* genome found below. For example, *B. thetaiotaomicron* contains 32 predicted family 2 glycoside hydrolases. (NC-not classified)

1.5 PUL75

A recent study of *B. thetaiotaomicron* showed that 13 of the 17 genes in PUL75 (Figure 1.7) showed an increased level of mRNA transcripts when the bacterium was grown in minimal media supplemented with HG⁷⁵. Of these 13 genes, seven are putative enzymes that belong to CE, GH and PL families associated with pectin utilization. These seven putative enzymes were identified based on sequence identity, as defined by the Carbohydrate-Active Enzyme Database (CAZy; www.cazy.org)⁴³. Among these seven enzymes, the five that belong to the GH and PL families (BT4108, BT4115, BT4116, BT4119 and BT4123) represent a significant gap in the understanding of PUL75's mode of action, as their substrate specificity and product profiles have yet to be determined. Interestingly, PUL75 also contains two copies of the genes responsible for regulation and sensing (HTCS in Red), and two of each the SusC- and SusD-like transporters (violet and yellow, respectively). The presence of two copies of the sensors (HTCS) and transporters (SusC- and SusD-like) suggests that there is redundancy, or selective regulation by a different inducer or substrate homolog. Finally, two copies of ORFs (*BT4112* and *BT4120*) are also present; however, their characterization and subsequent classification remains unknown.

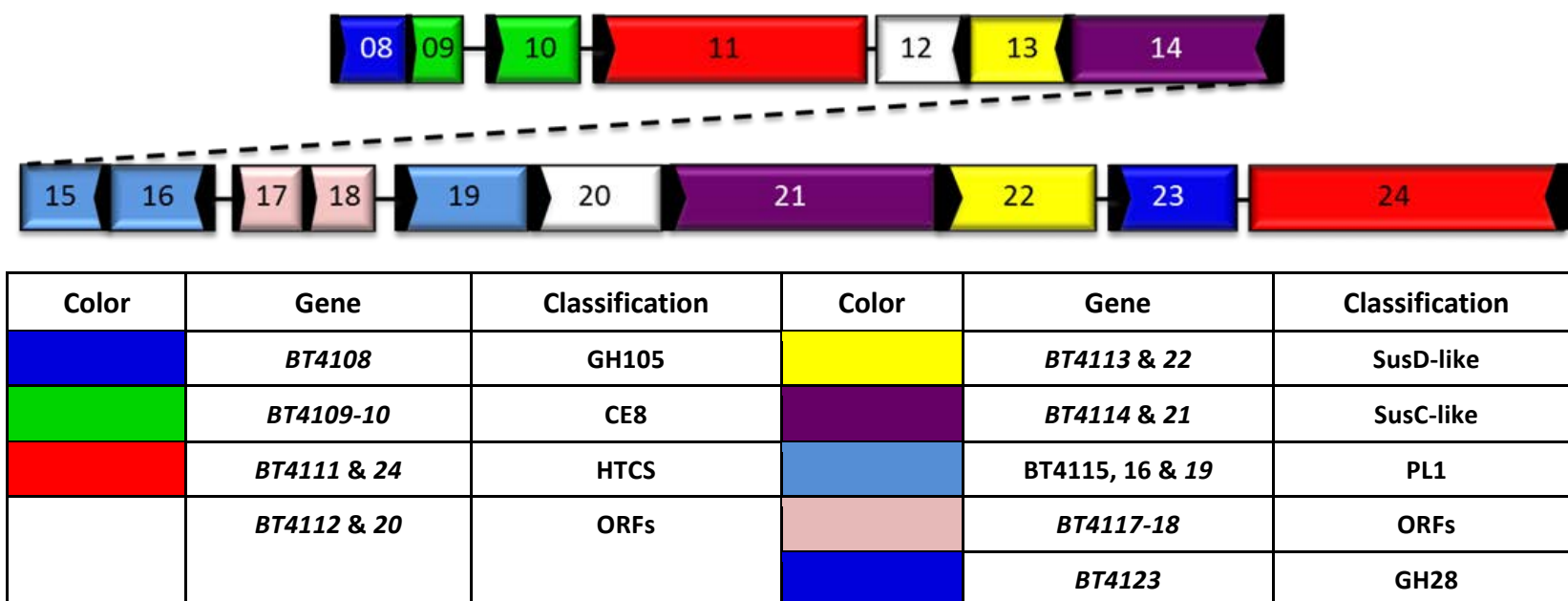


Figure 1.7: Visual representation of the seventeen genes which comprise PUL75. PUL75 contains two copies of the genes responsible for regulation and sensing (HTCS in Red), two of each the SusC and SusD -like transporters (violet and yellow, respectively) and three enzyme families; CEs (green), GHs (dark blue), and PLs (light blue). In addition, there are four open reading frames (ORFs, white and pink, respectively). Directionality is indicated by the black arrow and genes are drawn to scale. The black dashed lined indicates that these genes are contiguous. GH- Glycoside Hydrolase , CE- Carbohydrate Esterase, HTCS- Hybrid Two Component System, Sus- Starch Utilization System-like analogs, PL- Polysaccharide Lyase, ORFs- Open Reading Frames

Elucidating the systematic deconstruction of dietary fiber by unique members of the DGM may provide further insight into the complex relationship between human intestinal health, diet, and intestinal bacteria. Towards this end, a variety of molecular and biochemical approaches have been employed to characterize the function of *B. thetaiotaomicron* PUL75 gene products which are thought to be involved in HG modification, sensing and regulation and transport. Based on our characterization of PUL75 gene products, presented in the following chapters, we propose a model for the step-wise utilization of HG and have reconstituted an artificial HG deconstruction pathway that could have implications in biomass bioconversion and bioproduct generation.

Chapter 2-

Functional characterization of three putative family 1 polysaccharide lyases from PUL75

2.1 Introduction

Polysaccharide lyases (PL) are a class of enzymes that cleave glycosidic linkages via a β -elimination mechanism generating an unsaturated hexenuronic acid and a new free reducing end (uGalA_n). Within this larger classification, families of sequence related homologs have been defined that share similar folds and have conserved catalytic machinery (www.CAZy.org)⁴³. To date, 23 classified PL families exist, denoted by a numerical value following the PL (i.e. PL1). Of these 23 classified families, members of the PL1, PL2, PL3, PL4, PL9, PL10, PL11 and PL22 families have demonstrated activities on pectic polysaccharides. Structural characterization of these PL families has revealed a number of different folds including α/α -barrels, β -sandwich, β -helix, and β -propeller, and different requirements for metal co-factors. A brief outline of the mechanism, activities and structural highlights for HG active PLs follows.

2.1.1 PL mechanism— The HG substrate is an ideal candidate for β -elimination as the constituent GalA residues are in the 4C_1 conformation, with the C5 hydrogen and C4 hydroxyl group positioned in opposing axial configurations. The reaction progresses through an e1cb pathway (H-C α cleavage)⁸⁷, in which the rate-limiting step is C5 proton abstraction by a Brønsted base (Figure 2.1A). Proton acidification is facilitated by the C5

uronate, catalytic divalent metals, and localized basic residues within the active site, which draw charge from the C5 carbon and alter the local pKa environment. Most commonly the Brønsted base is a catalytic arginine (PLs 1, 2, 3, and 10) or lysine (family 9)^{57, 58}, which corresponds to an unusually high observed pH optimum.

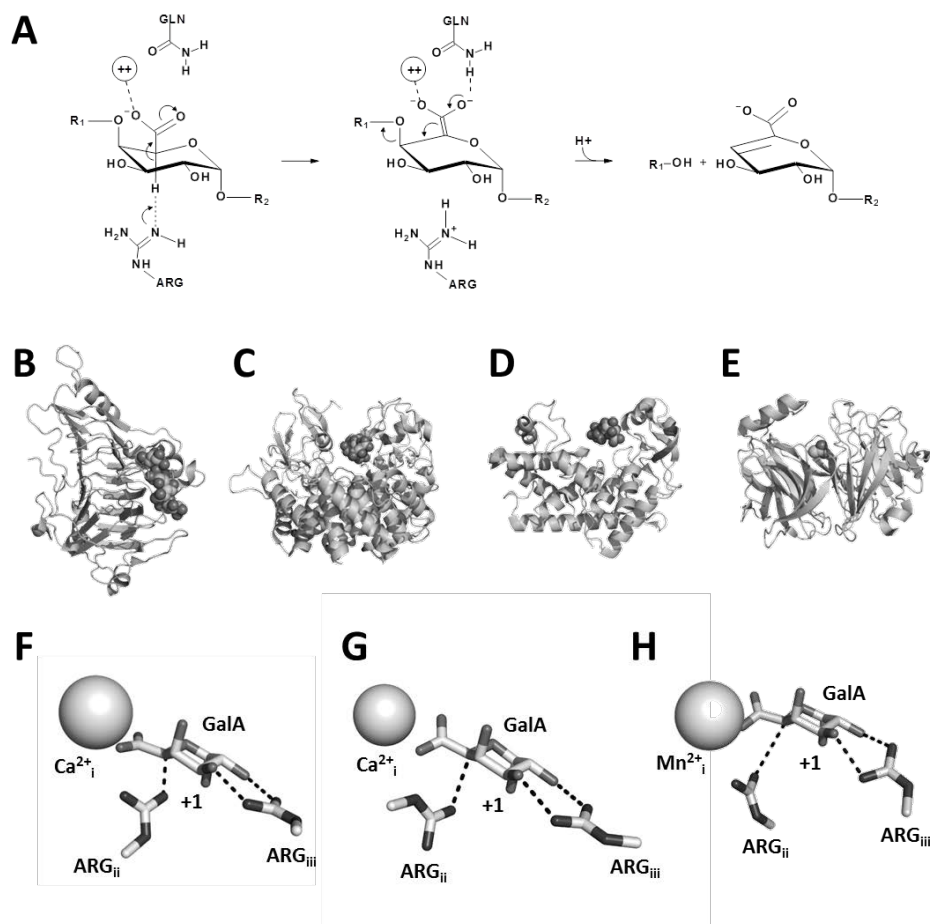


Figure 2.1: Structure and function of PLs active on HG. (A) β -elimination of α -D-GalA configured substrate. The metal cofactor is delineated as (++). (B-E) Three-dimensional structures of polysaccharide lyase-complexes active on HG and RG-I shown in cartoon representation with ligands as spheres. (B) PL1 β -helix from *E. chrysanthemi* EC16 in complex with GalA₆ (PDB ID: 2EWE)¹³. (C) PL2 α/α -barrel from *Y. enterocolitica* subsp. *enterocolitica* 8081 in complex with GalA₃ (PDB ID: 2V8K)⁸⁸. (D) PL10 α/α -barrel from *C. japonicus* Ueda107 in complex with GalA₃ (PDB ID: 1GXO). (E) PL22 β -propeller from *Y. enterocolitica* subsp. *enterocolitica* 8081 in complex with acetate (PDB ID: 3PE7). (F-H) Evolutionary convergence of the +1 subsite and β -elimination in pectate lyases⁸⁷⁻⁸⁹. The metal (i), Brønsted base (ii), and stabilizing Arg (iii) are shown for PL1 (F), PL9 (G), and PL2 (H) respectively.

This observation also reveals a dichotomy in PL function as the common pH optimum (8.0-9.5)⁹⁰ is more alkaline than the biological environments these enzymes are secreted into and are active in. The lowered optima for intracellular lyases suggest that this pH effect is primarily related to secreted PL function^{88, 89, 91}. The transition state proceeds through an enolate-enolate intermediate, which was recently reported to be resonance stabilized by hydrogen bonding or donation to the oxyanion from a dedicated lysine in family 1 and asparagine in family 9 PLs⁹². Decomposition of the intermediate occurs by protonation of the scissile glycosyl oxygen, through a yet to be determined mechanism, and elimination of the axial O4 creating an unsaturation between C4 and C5. The unsaturated product (uGalA), distorts the pyranosyl GalA conformation from a ⁴C₁ into a planar geometry, which is unstable for cyclized monosaccharide products.

2.1.2 PL activities— PL activities have been described as both exo and endo-acting enzymes⁴³. Exolytic PLs remove terminal sugar residues from the substrate, releasing shorter unsaturated oligosaccharides of a consistent size. Endolytic PLs cleave within the substrate, generating longer unsaturated oligosaccharides of differing sizes. Interestingly, PL families can contain both exo and endo-acting enzymes. This results in pectinolytic microorganisms often containing multiple copies of PLs from the same family within their genomes⁵⁸. In some cases, such as *Bacteroides* spp.⁴³, this redundancy is explained by tailored regulation of unique catabolic pathways (HG, RG-I, RG-II) or the existence of differential activities within a common family (exo and endo).

2.1.3 PL structural highlights— The first pectinase structure solved was PelC from *Erwinia chrysanthemi* EC16⁹³ (Figure 2.1B). PelC adopts a right-handed β -helix that

coils into three parallel β -sheets that are stabilized by aromatic stacks that run longitudinally through the protein core. At the time of its discovery, the structure of PelC defined a novel fold family. Since, it has proven to represent a plastic scaffold with utility for diverse pectinase activities⁹⁴, including several PL families (PL1, 3, and 9), and more surprisingly, distinct pectinase enzyme classes (PLs, GH28s, and CE8s)⁵³. The structural conservation in β -helix enzymes are believed to be a product of fold stability^{53, 94, 95}, as pectinases are commonly secreted into harsh and competitive environments, such as the gastrointestinal tract of animals, soil, and plant cell walls. More recently, several new fold families have been described for PLs including the PL2 α/α_7 -barrel (Figure 2.1C)⁸⁸, PL10 α/α_3 -barrel (Figure 2.1D)⁸⁷, and PL22 β 7-propeller (Figure 2.1E)⁸⁹. Despite this structural diversity, however, a common theme has emerged from the analysis of these enzymes. Investigation into the catalytic residues, metal cofactors, and substrates within active sites of these fold families has revealed a functional convergence of three key substructures (i-iii) that appear to be prerequisites for β -elimination (Figure 2.1F-H)^{89, 96}. These substructures include (i) a metal coordination pocket, (ii) a Brønsted base, and (iii) a stabilizing arginine. There is plasticity in two of these substructures as the metal binding pocket displays tailored chemistries specific for Ca^{2+} or different metals (e.g. Mn^{2+} and Mg^{2+})^{88, 89, 91} and the catalytic base has been determined to be most commonly an arginine, lysine⁹⁶ and perhaps histidine⁸⁹. The stabilizing arginine, however, is invariant which suggests that it may be essential for catalysis⁸⁹.

2.1.4 PUL75 PL1s— As previously described, PUL75 is an organized cluster of genes found within *B. thetaiotaomicron* that is thought to target HG metabolism⁷⁵. Within this

cluster, there are three predicted members of family 1 PLs; BT4115, BT4116 and BT4119. Though not uncommon for members of the same family to be present in the CAZome⁵⁸, the presence of these three family 1 PLs within the same PUL is interesting. Do all of these enzymes function as PLs? Do these PLs display unique preferences towards substrate and/or different activities, or are they redundant within PUL75? In light of this, we sought to clone and to produce recombinantly the three PL1s in order to facilitate their characterization, and elucidate their function in PUL75.

2.2 Materials and Methods

2.2.1 Recombinant protein production— The sequences coding the gene products of the three putative Polysaccharide Lyase family 1 (PL1) enzymes from PUL75, *BT4115*, *BT4116* and *BT4119*, were amplified from *B. thetaiotaomicron* genomic DNA using the indicated primers (Appendix 1). Amplicons were restricted with NheI (5') and XhoI (3') enzymes and ligated into pET28a (Novagen, Cat #69864-3) or NcoI (5') and XhoI (3') pET32a (Novagen, Cat #69015-3) expression vectors with complementary ends to create the expression plasmids pET28-BT4115, pET32-BT4116, and pET32-BT4119. These constructs contain N-terminal 6x histidine cleavable tags for purification by immobilized metal affinity chromatography (IMAC). All expression plasmids were individually transformed into BL21 (DE3) (EMD Millipore, Cat #69450-3, Karmstadt, Germany) cells for protein overexpression. These transformants were used to inoculate 1 l of LB broth containing kanamycin (50 $\mu\text{g ml}^{-1}$: pET28a) or ampicillin (50 $\mu\text{g } \mu\text{l}^{-1}$: pET32a). Cultures were grown at 37°C to an OD₆₀₀ of 0.8-1.0, cooled to 16°C and induced with 0.2 mM IPTG (Cells were continuously shaken at 180 rpm during the growth and induction process). The following day the cells were harvested by centrifugation at 6,500 x *g* for 10 min at 4°C. The cell pellet was re-suspended in sucrose solution [50 mM tris-HCl (pH 8.0), 25% sucrose with 10 mg l⁻¹ lysozyme] and chemically lysed at ambient temperature with 2 volumes of lysis buffer [1% deoxycholate, 1% triton X-100, 20 mM tris-HCl (pH 7.0), and 100 mM NaCl] with 200 $\mu\text{g l}^{-1}$ DNase. The lysate was clarified by centrifugation at 17,500 x *g* for 45 min at 4°C and filtered using a 0.45 μM low protein binding syringe filter (Pall Acrodisc 25mm). This soluble protein fraction was applied to a nickel (II) nitrilotriacetic acid (Ni²⁺-NTA) sepharose support (GE Healthcare, Cat#: 17-

5318-06) and the immobilized protein was washed with 500 mM NaCl, 20 mM tris-HCl (pH 8.5) to decrease non-specific binding. A stepwise imidazole gradient (5 mM-500 mM) was used to elute the target protein. Samples were visualized by SDS-PAGE, and fractions with appreciable amounts of target protein were pooled and dialyzed overnight against 4 l of 20 mM tris (pH 8.0), with two buffer exchanges, using a dialysis membrane (Spectrum, Biotech CE) with 5,000 MW cutoff. The following day, protein concentration was determined at OD_{280nm} using extinction coefficients of 70,750 M⁻¹ cm⁻¹ (1.248 mg ml⁻¹); 77,865 M⁻¹ cm⁻¹ (1.360 mg ml⁻¹); and 113,970 M⁻¹ cm⁻¹ (1.826 mg ml⁻¹), respectively for BT4115, BT4116 and BT4119, as calculated by the ProtParam analysis tool ⁹⁷. BT4119 was produced in the same manner as the two other PL1s; however, following lysis BT4119 was determined to be insoluble. The pellet was washed twice with distilled water, once with 20 mM tris-HCl (pH 8.0), 0.5% (v/v) triton X-100 and 500 mM NaCl, and twice with 20 mM tris-HCl (pH 8.0) and 500 mM NaCl. The washed pellet was dissolved in 6.0 M urea, 20 mM tris-HCl (pH 8.0), and 500 mM NaCl with vigorous stirring at ambient temperature over the weekend. The solubilized pellet was clarified at 20,000 x g for 45 min, filtered (0.45 µm), and applied to a Ni²⁺ - NTA column pre-equilibrated in 20 mM tris-HCl (pH 8.0), 6.0 M urea and 500 mM. Immobilized protein was washed with equilibration buffer, and a step-wise imidazole gradient (5-500 mM) was applied to elute the target protein. Eluted protein was dialyzed overnight in 2 l of 20 mM tris-HCl (pH 8.0), 500 mM NaCl, 2 M urea, 5% glycerol and 5 mM dithiothreitol (DTT) using a 5,000 MW cutoff dialysis membrane (Spectrum, Biotech CE). The following morning, the protein was further dialyzed against 2 l of 20 mM tris-HCl (pH8.0), 500 mM NaCl, 5% glycerol and 5 mM DTT for 3 h. A subsequent

round of dialysis in 20 mM tris-HCl (pH8.0), 250 mM NaCl, and 1% glycerol followed for 3 h. The protein was then finally dialyzed in 20 mM tris-HCl (pH 8.0) overnight, with two buffer exchanges. The following day the concentration of the pooled samples was determined at OD₂₈₀.

2.2.2 HG digestions— Overnight digests of HG were performed with each enzyme individually to determine substrate specificity and to generate product profiles. For each of the three PL1 family enzymes (BT4115, BT4116 and BT4119) reactions were incubated overnight at 37°C with 1 mg ml⁻¹ HG in 20 mM tris-HCl (pH 8.5). Calcium (0.1, 1.0 and 2.0 mM: BT4115, BT4119 and BT4116, respectively) was chelated with EDTA (5 mM) to investigate metal dependence on PL1 activity. After incubation the samples were heat treated at 100°C for 10 min to denature the enzyme and terminate the reaction. The samples were centrifuged to remove precipitate and spotted onto a silica plate with aluminum backing for thin layer chromatography (TLC; EMD, Cat #55553-7, Karmstadt, Germany). Galacturonic acid (GalA, Sigma Cat# 48280, St. Louis, MO), digalacturonic acid (GalA₂, Sigma, Cat# D4288, St. Louis, MO) and trigalacturonic acid (GalA₃, Sigma, Cat# T7407, St. Louis, MO) were used as standards. The plates were run twice in a mobilization buffer consisting of 1-butanol, distilled water and acetic acid (5:3:2) and stained with 1% orcinol in a solution of ethanol and sulfuric acid (70:3 v/v). The plate was incubated at 120°C for 2-5 min in a dry oven for visualization.

2.2.3 PL1 HG digestion time courses— After initial activity screening, time courses were prepared for each of the three PL1 enzymes (BT4115, BT4116 and BT4119). Individual time courses were prepared with final concentrations of 1 mg ml⁻¹ HG (Megazyme), 20

mM tris-HCl (pH 8.5), calcium (see 2.2.2) and 1 μ M of the respective enzyme. The master mix was incubated at 37°C, and samples taken at 0, 5, 10, 15, 30, 45, 60, 120, 180, 300 min, 1, 2, and 4 days. After a given time point was taken, the sample was heat treated at 100°C for 10 min. Due to the inability to visualize digestion products, each individual time point for BT4116 and BT4119 HG digestion was lyophilized, rehydrated, centrifuged and spotted on the TLC. Depending on the size of products expected, the TLCs were run a minimum of twice in mobilization buffer (as above), however BT4119 was run three times for maximum resolution of products. Plates were stained with 1% orcinol in a solution of ethanol and sulfuric acid (70:3 v/v) and incubated at 120°C for 2-5 min in a dry oven for visualization.

2.2.4 High performance anion-exchange chromatography with pulsed amperometric detection (HPAEC-PAD) product analysis— HPAEC-PAD runs were performed on an ICS DIONEX 3000 system using a CarboPac PA-20 column (DIONEX, Cat #060142, Sunnyvale, CA) and a 0 – 900 mM sodium acetate (NaOAc) gradient in 100 mM NaOH with a 0.5 ml min⁻¹ flow rate. Samples were boiled, centrifuged, frozen prior to analysis, and diluted 1:5-1:20 (250 μ l final volume) before loading. Retention times were determined and compared to GalA, GalA₂ and GalA₃ standards.

2.2.5 PL1 enzyme matrix digestions— An enzyme matrix (Table 2.1) was created to investigate complementarity between the three PL1s. Pilot digestions of each individual enzyme were carried out to determine relative enzyme concentrations to be used in the enzyme matrix (0.3 μ M BT4115 and 1.0 μ M BT4116 and BT4119). The seven different reactions (Table 2.1) were set up under the determined digestion conditions (20 mM tris

pH 8.5, 1 μM CaCl_2 and 1 mg ml^{-1} HG) and aliquots taken at 1, 6, 12 and 24 h. These aliquots were heat treated at 100 $^{\circ}\text{C}$ for 10 min and analyzed via TLC and HPAEC-PAD, as described previously. The area under the resulting product peaks was calculated using the ChromeleonTM software to quantify the amount of products produced.

Table 2.1: Matrix digestion template to investigate PL1 complementary activities.

Reaction	i	ii	iii	Iv	v	vi	vi
	<i>Enzymes added</i>						
BT4115	X	X		X	X		
BT4116	X	X	X			X	
BT4119	X		X	X			X

2.2.6 Spectrophotometric enzyme assays— All enzyme assays with recombinant BT4115 were performed at 0.1 μM , with 20 mM tris (pH 8.5) and 0.1 mM CaCl_2 . Unsaturated oligogalacturonides were detected using UV spectrophotometry at 232 nm, while incubated at 37 $^{\circ}\text{C}$, to quantify the generation of double bonds [Extinction Coefficient (Ext Co) = 4,600 $\text{M}^{-1} \text{cm}^{-1}$]⁹⁸. The total reaction volume was 600 μL in quartz cuvette. The linear portion of the reaction curve was used to determine the rate of production. Data points were collected in triplicate, from three different consecutive runs, and processed with Graph Pad Prism 6TM.

2.2.7 Spectrophotometric enzyme assays using divalent cations— Soluble BT4115 was prepared as above. Prior to the final dialysis step soluble BT4115 was dialyzed against 20 mM tris (pH 8.0) and 1 mM EDTA to chelate any divalent cation present. Following chelation, BT4115 was aliquoted into seven equal parts. Six were then dialyzed twice against 20 mM tris (pH 8.0) and 1 mM divalent cation from calcium (Ca^{2+}), magnesium

(Mg²⁺), manganese (Mn²⁺), cobalt (Co²⁺), nickel (Ni²⁺) and zinc (Zn²⁺) while the seventh was dialyzed twice against 20 mM tris (pH 8.0). Following the two rounds of dialysis, protein concentration was determined at OD₂₈₀ using an extinction coefficient of 70,750 M⁻¹ cm⁻¹ (1.248 mg ml⁻¹). The enzymatic analysis was performed with each BT4115 sample at 0.1 μM, with 20 mM tris (pH 8.5) and 0.1 mM corresponding divalent cation. Products resulting from the digestion of varying concentrations of HG (Sigma, Cat # 81325, St. Louis, MO) were detected as in section 2.2.6. When BT4115 was dialyzed against Zn²⁺, the soluble BT4115 precipitated, resulting in a loss of all active BT4115, thus the kinetic analysis was not performed. A TLC plate was used to visualize 1 and 24 hr digestions of HG with BT4115 and a select divalent cation (Ca²⁺, Mg²⁺, Mn²⁺, Co²⁺ and Ni²⁺). Digestions were incubated at 37°C with final concentrations of 1 μM enzyme, 20 mM tris (pH 8.0) and 0.1 mM of divalent cation. Time points were heat treated at 100°C for 10 min to terminate the reaction. Samples were centrifuged at high speed and the supernatant spotted onto the TLC and visualized as in section 2.2.2.

2.2.8 Oligosaccharide production, separation and purification— Small scale digestions of HG with BT4115 were performed as in section 2.2.2 to target unsaturated oligosaccharides (2-7 sugars in length). Incubation times of the small scale digests were adjusted to produce unsaturated oligosaccharides 2-7 sugars in length. Digestions were subsequently scaled up 500 times to produce large quantities of these unsaturated oligosaccharides. Once cooled, the samples were flash frozen using liquid nitrogen and lyophilized. The lyophilized digestion products were re-suspended in approximately 1 ml distilled water and centrifuged at 14,000 rpm for 5 min. The supernatant was loaded onto a column with P-2 resin (Bio-Rad) equilibrated in 0.1 M acetate buffer (pH 3.6) with a

flow rate of 0.25 ml min⁻¹ (Pharmacia P-1 peristaltic pump). Column fractions were collected (Bio-Rad Fraction Collector Model 2110) at 9 minute intervals. Collected fractions were screened for the presence of oligosaccharides by TLC. Target oligosaccharides were further analyzed via TLC and fractions pooled based on length, flash frozen and lyophilized. Lyophilized products were then re-suspended in ~1 ml and loaded onto a column with P-2 resin equilibrated in distilled water with a flow rate of 0.25 ml min⁻¹. Column fractions were collected at 9 minute intervals, and collected fractions were screened for the presence of oligosaccharides by TLC. Target oligosaccharides were further analyzed by TLC and fractions pooled based on length, flash frozen and lyophilized. The lyophilized product (powder) was weighed and stored at -20°C.

2.3 Results

2.3.1 BT4115

2.3.1.1 Cloning, protein expression and purification— The mature gene sequence of BT4115 was cloned into the pET28a *E. coli* expression vector in order to produce recombinant protein for characterization. Gene sequencing confirmed no mutations in the construct. BT4115 was produced as an approximately 56 kDa soluble protein and purified by immobilized metal affinity chromatography (IMAC). BT4115 fractions were pooled based on SDS-PAGE analysis (Figure 2.2D) and dialyzed. Approximately 20 mg of BT4115 was obtained from 1L culture as determined by A_{280} .

2.3.1.2 pH and CaCl_2 optimum— The pH optimum of BT4115 was investigated by Erin Crawford (former Co-op student) using a buffer range from pH 6.8-10.4 by detecting the unsaturated oligogalacturonide products at 232 nm (Figure 2.3). From these analyses the pH range of 7.5-9.5 displayed optimal activity. The optimal CaCl_2 concentration was also investigated with 0.1 mM displaying optimal activity (data not shown). Based on the optimum range obtained, all subsequent activity assays involving BT4115 utilized the following conditions; 20 mM tris pH 8.5, 0.1mM CaCl_2 .

2.3.1.3 Ca^{2+} dependence and kinetic analyses— BT4115 metal dependence was investigated by EDTA treatment of the enzyme to chelate Ca^{2+} ions. Digestion of HG in the presence of EDTA revealed a loss of activity (Figure 2.4A). Further investigation into Ca^{2+} cation preference was determined by chelating and dialyzing the enzyme with different divalent cations [Mg^{2+} , Mn^{2+} , Cu^{2+} , Ni^{2+} and Zn^{2+}] and performing kinetic

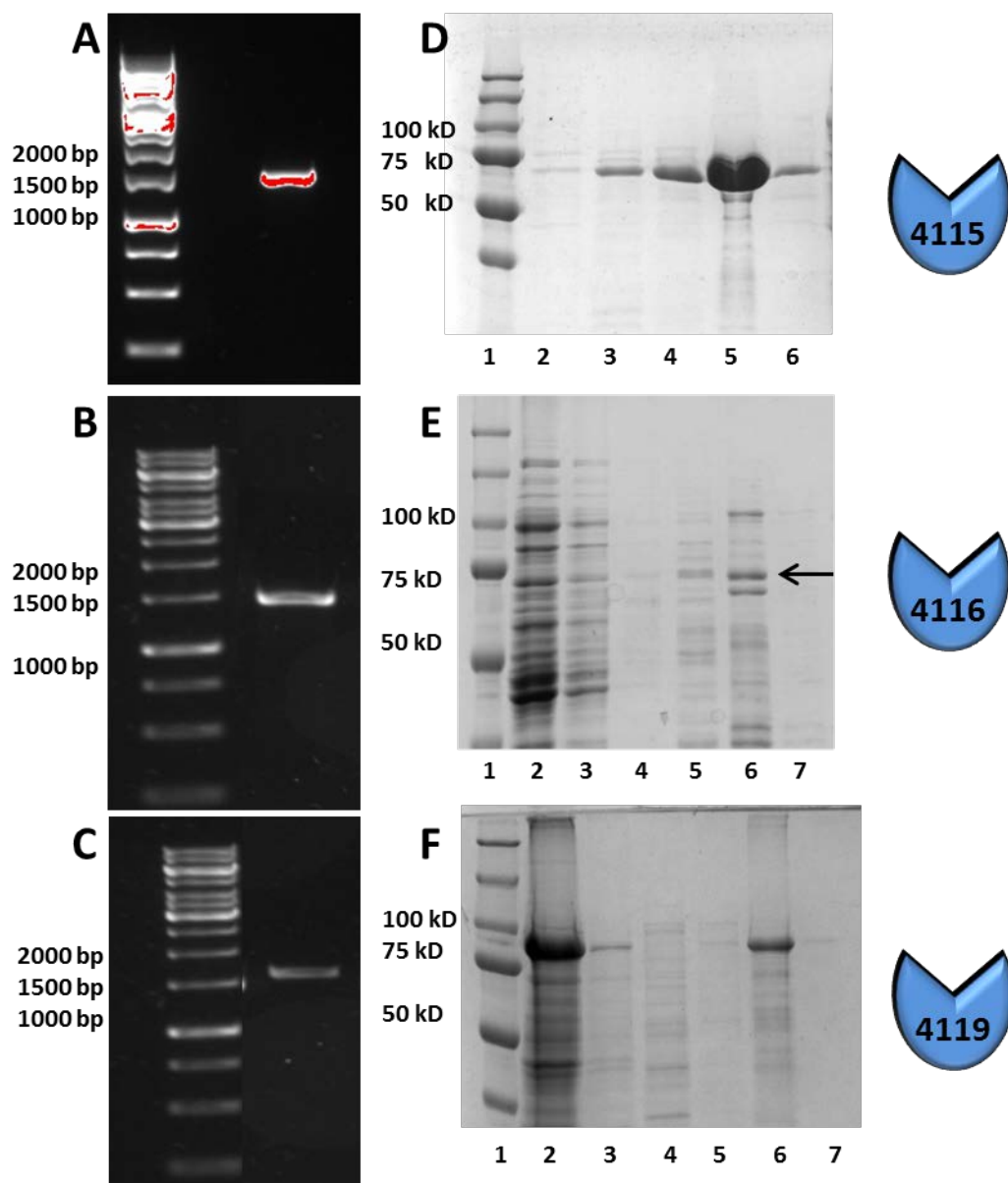


Figure 2.2: Cloning and recombinant protein purification of the three PL1s from PUL75. (A, B and C) Agarose Gel (1%) displaying the PCR product for each of the respective PL1 genes, *BT4115*, *BT4116* and *BT4119*, respectively. The MWM is on the left hand side of the gel, and the PCR product is on the right. (D) SDS-PAGE analysis of *BT4115* (pET28a) IMAC fractions. Lanes 1-6; MWM, Wash, 5 mM, 10 mM, 100 mM and 500 mM elution, respectively. (E) SDS-PAGE analysis of *BT4116* (pET32a) IMAC fractions. Lanes 1-6; MWM, cell lysate, Wash, 5 mM, 10 mM, 100 mM and 500 mM imidazole, respectively. (Arrow indicates *BT4116* on gel). (F)- SDS-PAGE analysis of *BT4119* (pET32a) IMAC fractions. Lanes 1-7; MWM, cell lysate, wash, 5 mM, 10 mM, 100 mM and 500 mM elution, respectively. Far right are cartoon representations of the three PL1 enzymes.

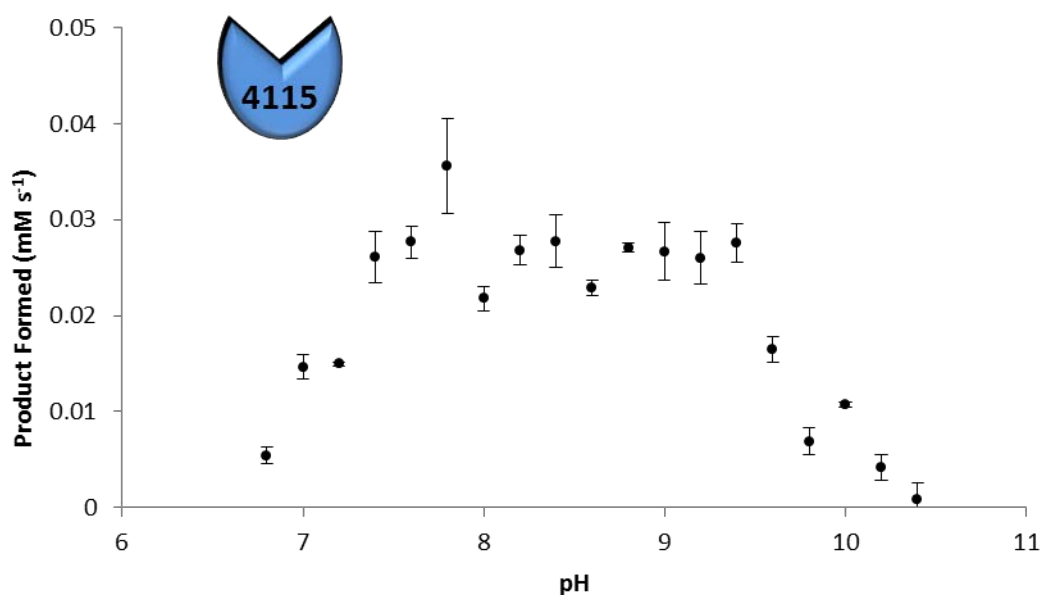


Figure 2.3: BT4115 pH optimum. Data points represent the generation of unsaturated product as determined by UV absorbance at 232 nm. Reactions were performed in triplicate, with the average taken and plotted. The *error bars* represent the standard deviation from the average.

analyses in the presence of the different metals (Figure 2.5A). The complete kinetic parameters calculated can be found in Table 2.2. Interestingly, when comparing the catalytic efficiencies (k_{cat}/K_m), both Ca^{2+} and Mg^{2+} displayed an approximately 10 to 100-fold increase compared to the remaining cations. Five previously characterized PLs from *Erwinia chrysanthemi* displayed a range of K_m values from 0.03-0.30 mg ml⁻¹ and a range of V_{max} values between 46 and 3800 $\mu\text{mol min}^{-1} \text{mg}^{-1}$.

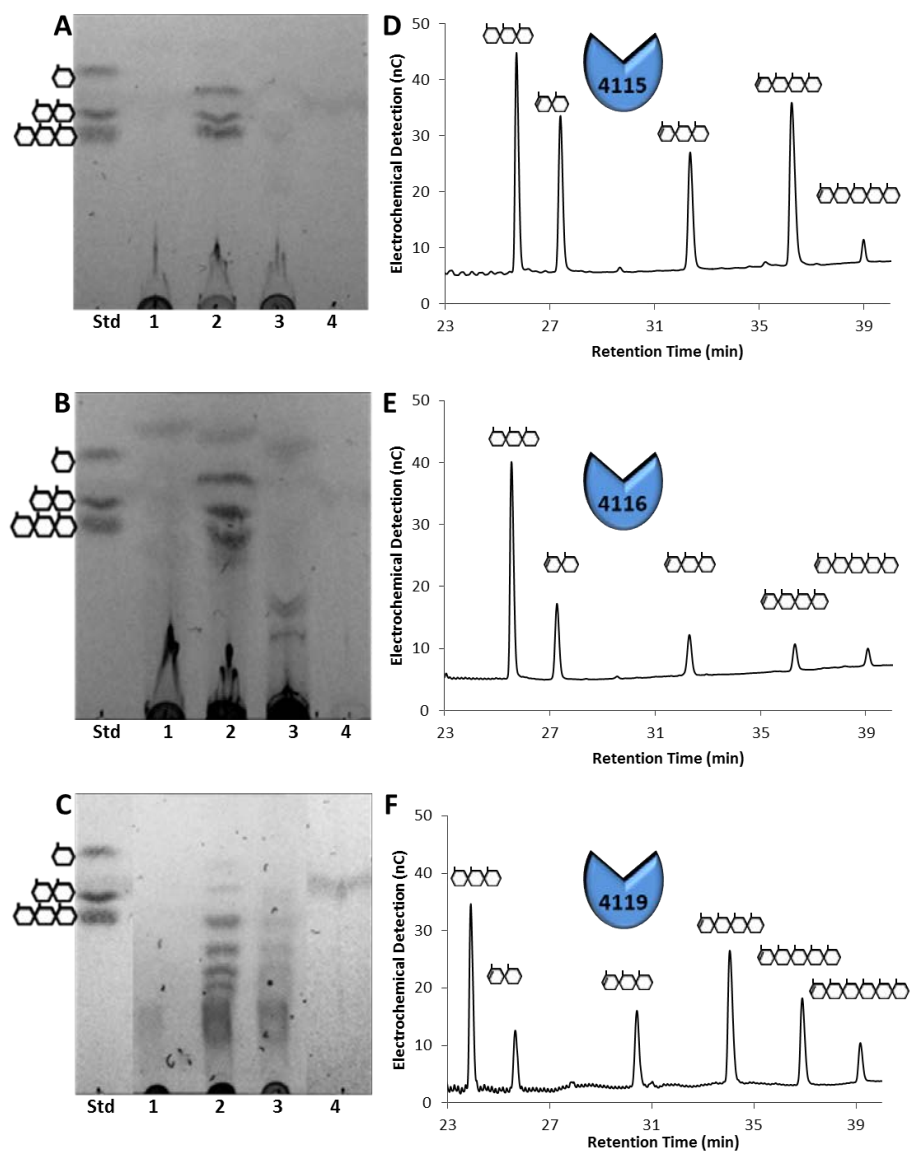


Figure 2.4: PL1 HG digestion product profiles (A-C) TLC analysis of BT4115, BT4116 and BT4119 HG digestion, respectively. Lanes: Std- mono-, di- and tri-galacturonic acid standards with sizes indicated on the left. 1-4; HG, HG and enzyme, HG + enzyme + EDTA, enzyme, respectively. **(D-F)** HPAEC-PAD analysis of overnight BT4115, BT4116 and BT4119 HG digestions, respectively.

Table 2.2: Kinetic parameters obtained from BT4115 HG digestions with divalent cations

Metal	$V_{\max}$ ($\mu\text{M min}^{-1}$)	K_m (mg ml^{-1})	K_{cat} (min^{-1})	$V_{\max}$ ($\mu\text{mol min}^{-1} \text{mg}^{-1}$)	K_{cat}/K_m (min^{-1})/ (mg ml^{-1})
Calcium	380 ± 28	0.19 ± 0.04	3800 ± 280	670 ± 50	20000 ± 5700
Magnesium	180 ± 7.0	0.08 ± 0.01	1800 ± 70	320 ± 12	23000 ± 3800
Manganese	72 ± 11	0.46 ± 0.02	720 ± 110	130 ± 20	1600 ± 300
Cobalt	29 ± 14	0.83 ± 0.72	290 ± 140	52 ± 25	350 ± 300
Nickel	110 ± 29	4.2 ± 1.5	1100 ± 290	190 ± 50	260 ± 160
<i>Calcium (supplemented)</i>	120 ± 11	0.033 ± 0.013	1200 ± 110	220 ± 20	37000 ± 18000
Zinc	N/A	N/A	N/A	N/A	N/A

2.3.1.4 Product profile— The products of BT4115 HG digestion were analyzed via TLC, HPAEC-PAD and mass spectrometry. Mass spectrometry was carried out at the Complex Carbohydrate Research Center (CCRC) in Athens, Georgia by Dr. Sami Tuomivaara. Mass spectrometry analysis confirmed the assignment of products visualized post digestion on TLC and with HPAEC-PAD (Figure 2.16). The results indicated that upon complete digestion, BT4115 produces uGalA₂, uGalA₃ and uGalA₄ (Figure 2.4D). A time course of BT4115 digestion revealed the generation of uGalA₅ and uGalA₆; however, these products were subsequently digested into smaller products as the reaction progressed (Figure 2.7A).

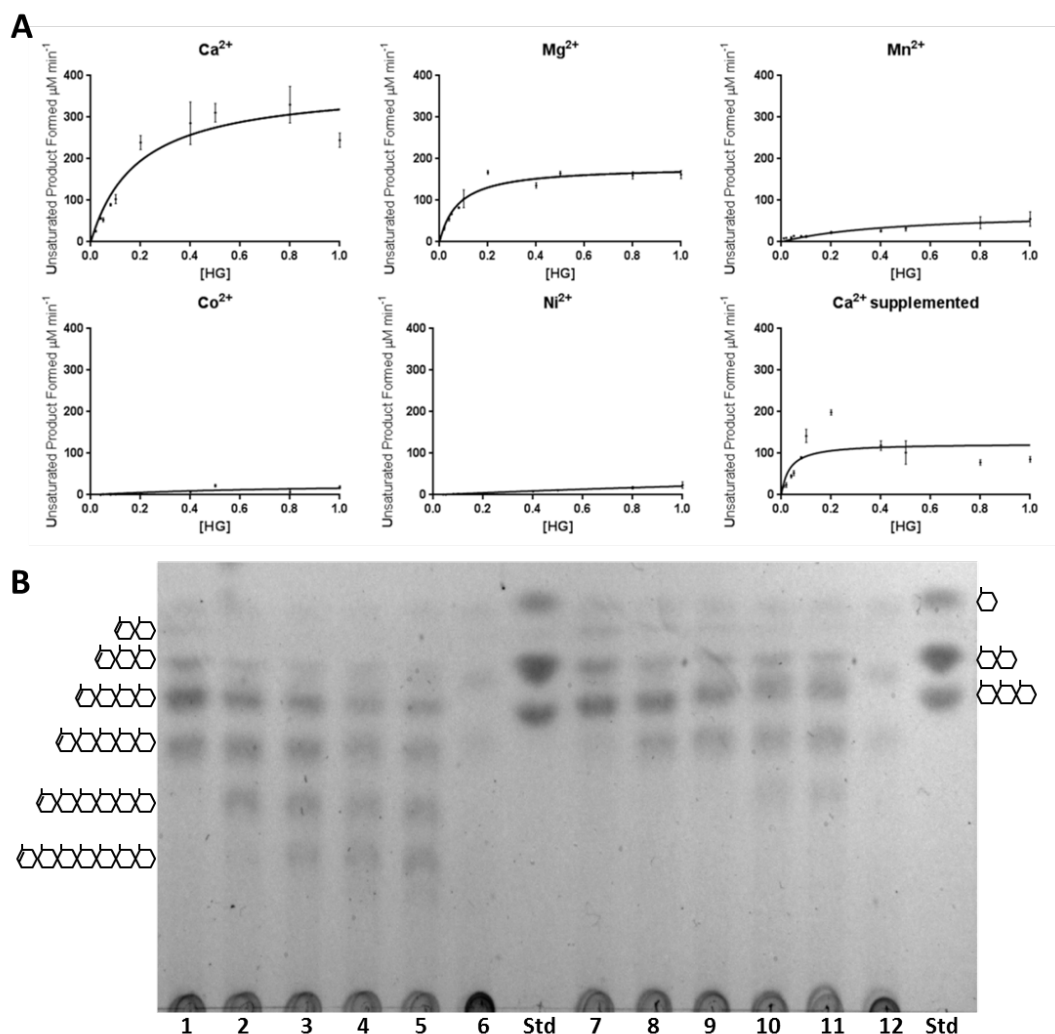


Figure 2.5: Kinetic analysis and product profiles of BT4115 with select divalent cations. (A) Graphs representing the kinetic analysis of the rate of BT4115 HG digestion with select divalent cations as indicated on the individual graphs. Ca²⁺, Mg²⁺, Mn²⁺, Ni²⁺, Co²⁺ and Ca²⁺ supplemented. All samples were prepared by chelating metals with EDTA, and subsequent dialysis with target divalent cation at saturating conditions. Calcium supplemented sample was not chelated with EDTA, or dialyzed with calcium, however calcium was added during kinetic analysis. All points were run in triplicate with the *error bars* representing the standard deviation. Data was analyzed and processed using the Graph Pad Prism 6™ software. (B) Time course of HG digestion by BT4115 and a select divalent cation. Samples correspond to those mentioned in (A). Lanes 1-6 represent 1-h time point of BT4115 and Ca²⁺, Mg²⁺, Mn²⁺, Ni²⁺, Co²⁺ and no divalent cation, respectively. Lanes 7-12 represent overnight digestion of BT4115 and Ca²⁺, Mg²⁺, Mn²⁺, Ni²⁺, Co²⁺ and no divalent cation, respectively. Std represents, GalA, GalA₂ and GalA₃ standards, sizes are indicated in the right hand margin. Sizes of products can be obtained by the left hand margin reference oligogalacturonides.

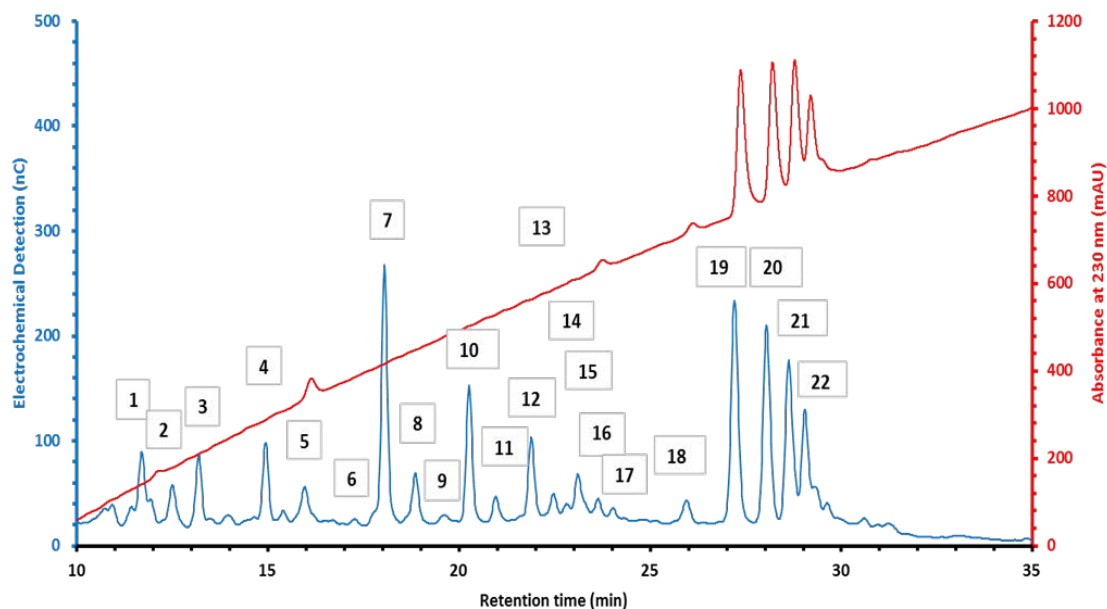


Figure 2.6: HPAEC-PAD analysis of large scale PL1 HG digestion. Large scale PL1 HG digest were analyzed via HPAEC-PAD. All sample peaks were collected on-line and subsequently analyzed via mass spectrometry to confirm identity. Left hand axis represents the reducing end detection of all analytes. Right hand axis represents the detection of unsaturated sugar moieties in the analyte. Each peak is identified by a corresponding number with further information provided in Table 2.3. (For mass spectrometry chromatograms see appendices 2-4)

Table 2.3: Mass spectrometry designation of peaks collected from HPAEC-PAD analysis

Peak #	Time	MALDI	m/z	ID	Peak #	Time	MALDI	m/z	ID
1	11.7	no		GalA	12	21.9	yes	921	GalA ₅
2	12.5	no			13	22.4	no		
3	13.3	no			14	22.8	yes	1097	GalA ₆
4	14.9	no		GalA ₂	15	23.1	no		
5	16	no		uGalA	16	23.6	no		uGalA ₂
6	17.3	no			17	24	no		
7	18	yes	569	GalA ₃	18	25.9	yes	551	uGalA ₃
8	18.8	no			19	27.2	yes	727	uGalA ₄
9	19.6	no			20	28	yes	903	uGalA ₅
10	20.3	yes	745	GalA ₄	21	28.6	yes	1079	uGalA ₆
11	21	no			22	29.1	yes	1255	uGalA ₇

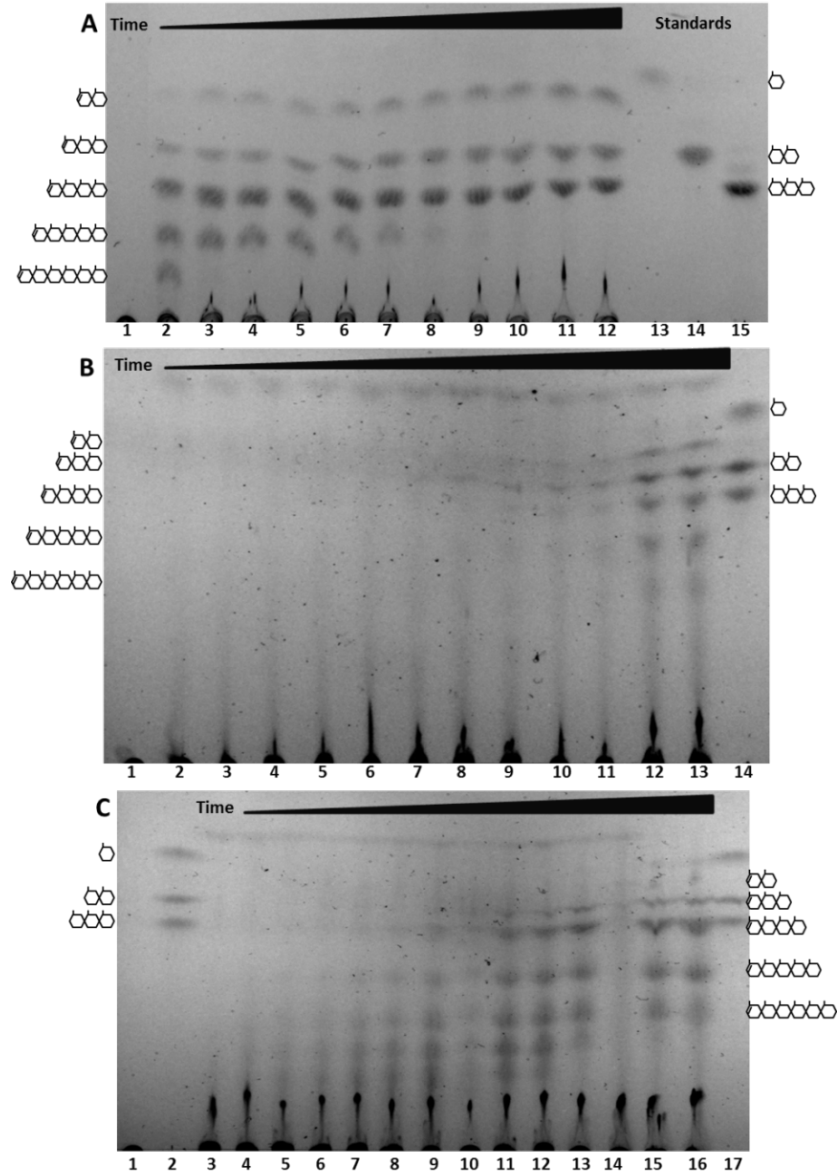


Figure 2.7: Time courses of HG digestions by the PUL75 PL1s BT4115, BT4116 and BT4119 (A) BT4115 HG digestion time course. Lanes 1-12, reaction time points (0,1,5,10,15,30,60,120,180 and 300 min, overnight and 24 h) Lanes 13-15, standards, GalA, GalA₂ and GalA₃, respectively. Sizes of products and standards are indicated in the margins on the left and right, respectively. **(B)** BT4116 HG digestions time course. Lanes 1-13, reaction time points (0, 5, 10, 15, 30, 60, 120, 180 and 300 min, overnight and 1-, 2- and 4-day(s)) Lane 14, standards, GalA, GalA₂ and GalA₃. Sizes of products and standards are indicated in the margins on the left and right, respectively. **(C)** BT4119 HG digestion time course. Lanes 1, 3-16, reaction time points (0, 5, 10, 15, 30, 45, 60, 120, 180 and 300 min, overnight and 1-, 2- and 4-day(s)).

2.3.1.5 Large scale digests— BT4115 was utilized to generate preparative quantities (mg) of unsaturated oligosaccharides (uGalA₂-uGalA₇) that are not commercially available. These products (Figure 2.8) were purified via chromatography, pooled based on degree of polymerization and lyophilized. The generated products were subject to mass-spectrometry analysis to confirm identity.

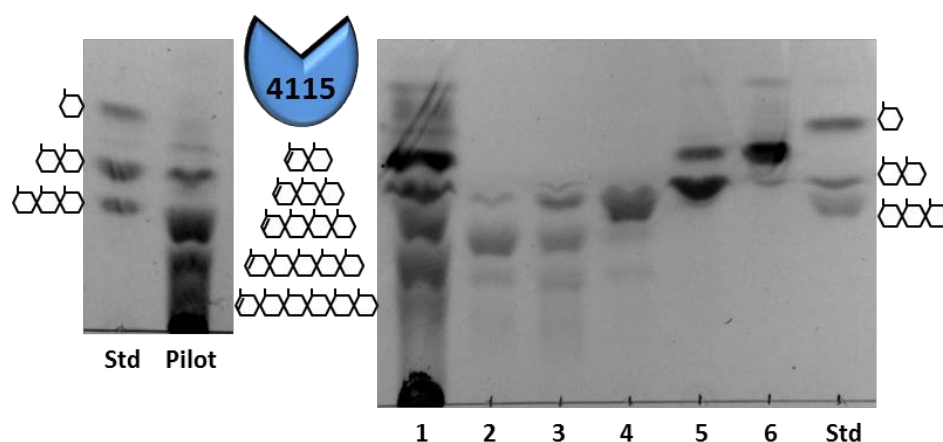


Figure 2.8: Large scale BT4115 HG digestions for oligogalacturonide generation. **(Left)** TLC analysis of pilot digestion. Std- GalA, GalA₂ and GalA₃ standards. Pilot denotes pilot digestion. **(Right)** TLC analysis of chromatography separated oligogalacturonide fractions. Lanes 1 digestions samples (pre-separation) lanes 2-6, separated oligogalacturonides. Product peaks are identified by corresponding carbohydrate in the margins.

2.3.2 BT4116

2.3.2.1 Cloning, protein expression and purification— BT4116 was successfully cloned into the pET32a *E. coli* expression vector (Novagen) which produced a soluble protein with a mass of approximately 70 kDa (due to inclusion of the thioredoxin fusion). The thioredoxin fusion was not cleaved, and all subsequent analysis occurred with the fusion protein, referred to as BT4116. Gene sequencing confirmed no point mutations in the gene construct. Soluble recombinant protein was purified by IMAC (Figure 2.2E).

BT4116 was pooled based on SDS-PAGE analysis (Figure 2.2E) and dialyzed. Approximately 2 mg BT4116 was obtained from 1L culture.

2.3.2.2 Ca²⁺ dependence and optimum— BT4116 metal dependence was investigated by EDTA treatment of the enzyme to chelate Ca²⁺ ions. Subsequent digestion with this chelated enzyme revealed a loss of activity (Figure 2.6B). The Ca²⁺ optimum of BT4116 was investigated spectrophotometrically using a range of concentrations by the detection of an unsaturated oligogalacturonide product at 232 nm (data not shown). From these analyses the optimal CaCl₂ concentration was found to be 2.0 mM, with 20 mM tris pH 8.5.

2.3.2.3 Product profile— The products of BT4116 HG digestion were analyzed via TLC, HPAEC-PAD and mass spectrometry. The results indicated that, upon completion, BT4116 produces uGalA₂, uGalA₃ and uGalA₄ (Figure 2.4E). A time course of BT4116 digestion revealed the generation of unsaturated oligogalacturonides, consistent with the overnight digestions. However, these products occurred much later in the reaction time course (Figure 2.7B).

2.3.3 BT4119

2.3.3.1 Cloning, protein expression and purification— The mature gene sequence of BT4119 was cloned into *E. coli* expression vectors to produce recombinant protein. Initially, the coding region of BT4119 was cloned into the pET28a vector (Novagen) by Erin Crawford, however, no soluble protein was obtained. BT4119 was subsequently cloned into the pET32a vector and again no soluble protein was obtained. An appreciable amount of the protein was found to be present in the insoluble fraction. The inclusion

bodies were disrupted and an approximately 80 kDa thioredoxin fusion target protein was isolated and purified by IMAC and re-folded by multiple rounds of dialysis. The re-folded BT4119 was pooled based on SDS-PAGE analysis (Figure 2.2F) and dialyzed exhaustively. Gene sequencing confirmed no mutations in the construct. Approximately 4 mg of BT4119 was obtained from 1L culture (insoluble preparation). The thioredoxin fusion was not cleaved, and all subsequent analysis occurred with the fusion protein, referred to as BT4119.

2.3.3.2 Ca^{2+} dependence— BT4119 metal dependence was investigated by EDTA treatment of the enzyme to chelate Ca^{2+} ions. Subsequent digestion with this chelated enzyme revealed a loss of activity (Figure 2.4C). Spectrophotometric analysis determined an optimal Ca^{2+} concentration of 1 mM in 20 mM tris pH 8.5 (data not shown)

2.3.3.3 Product profile— The products of BT4119 HG digestion were analyzed via TLC, HPAEC-PAD and mass spectrometry. The results indicated that upon completion, BT4119 produces uGalA₂ – uGalA₆ (Figure 2.4F). A time course of BT4119 digestion revealed the generation of larger unsaturated oligogalacturonides; however, these products were subsequently digested into smaller products as the reaction progressed (Figure 2.7C).

2.3.4 Complementary activities between PUL75's PL1s and PL1 Matrix

2.3.4.1 Complementary digestions— BT4115 was incubated with unsaturated oligosaccharides from BT4119 activity on HG (uGalA₂ – uGalA₆; Figure 2.9). The results of these digests, visualized by TLC and analyzed by HPAEC-PAD, indicate that

BT4115 digests the larger unsaturated oligosaccharide of the BT4119 product profile (Figure 2.9A, Lanes 3 & 4), resulting in a product profile identical to that of BT4115 HG digestion. The loss of larger unsaturated oligogalacturonides (from BT4119 profile) indicates the presence of complementary activities between BT4115 and BT4119.

BT4116 was incubated with the products of BT4115 digests of HG (uGalA₂, uGalA₃ and uGalA₄; Figure 2.9). The results of these digests, visualized by TLC, indicate that BT4116 further modifies uGalA₄ (Figure 2.9A, Lanes 1 & 2), resulting in a product profile consisting of a mixture of uGalA₂ and uGalA₃ (Figure 2.9A-B). These digestions were subsequently analyzed via HPAEC-PAD, which indicated an increase in the uGalA₂ peak. This corresponds to the conversion of uGalA₄ into two uGalA₂.

2.3.4.2 PL1 matrix digestions— A digestion matrix involving the three separate PL1 enzymes was set up according to Table 2.1. Common time points of each of the individual digests were obtained and analyzed via TLC and HPAEC-PAD (Figure 2.10-12). The area under the peaks for each of the analyzed products was calculated using the Chromeleon™ software. These quantitative results indicated that, as the length of digestion increases, there is a visible shift to shorter unsaturated oligosaccharides (Figure 2.13-15). Furthermore, an increase in the relative amount of these shorter length oligosaccharides is observed as the longer length oligosaccharides are consumed (Figure 2.13-15).

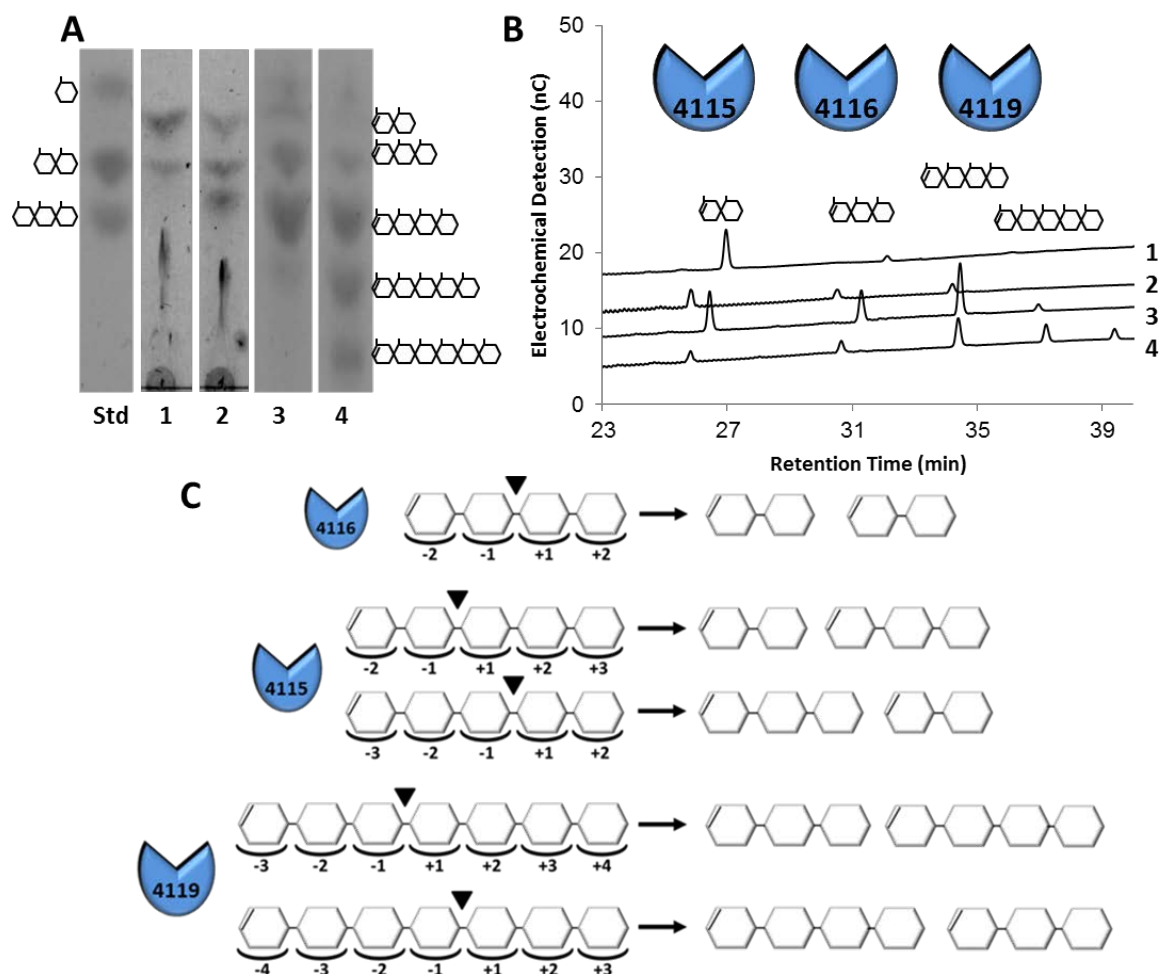


Figure 2.9: TLC and HPAEC-PAD analysis of PL1 complementary digestions. (A) TLC analysis of BT4115, BT4116 and BT4119 complementary digestions. Lanes: Std- GalA, GalA₂ and GalA₃ standards with sizes indicated on the left. Lanes 1-4; BT4116 + BT4115 products, BT4115 products, BT4115 + BT4119 products, BT4119 products. **(B)** HPAEC-PAD analysis of overnight BT4115, BT4116 and BT4119 complementary digestions. Chromatogram traces have a number corresponding to the TLC lane of the analyzed sample. Product peaks are identified by corresponding carbohydrate. **(C)** Proposed subsites for endolytic PL1s from PUL75.

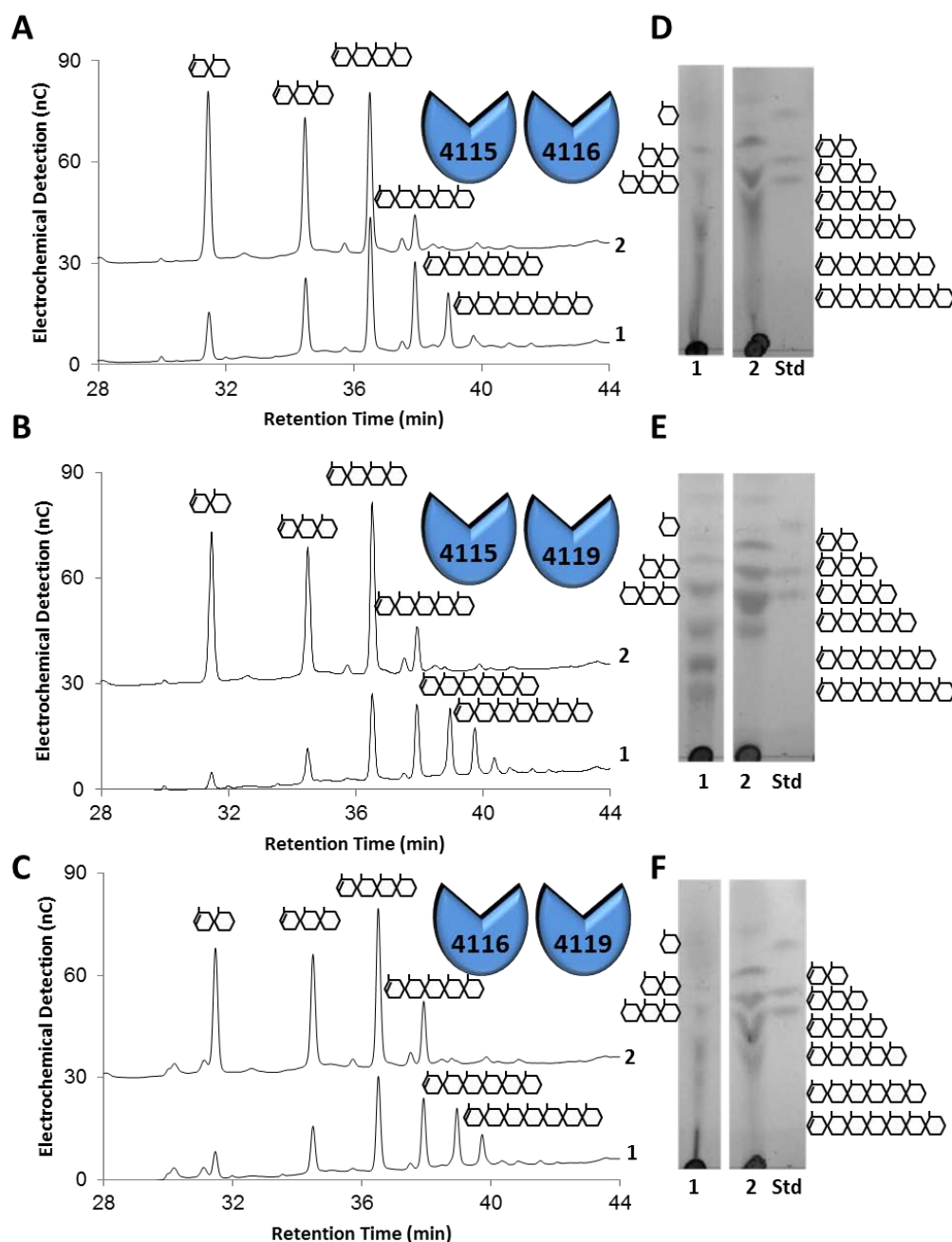


Figure 2.11: HPAEC-PAD and TLC analyses of PL1 BT4115, BT4116 and BT4119 HG time course digestions (A-C) HPAEC-PAD analysis of BT4115, BT4116 and BT4119 HG co-digestions. Co-digestion partners are indicated by cartoon representations of enzymes. (D-F) TLC analysis of corresponding BT4115, BT4116 and BT4119 HG co-digestions. Lanes 1 and 2, reaction time points (1 and 24 h, respectively). Std- GalA, GalA₂ and GalA₃ standards. Product peaks and standards are identified by corresponding carbohydrates in the margins.

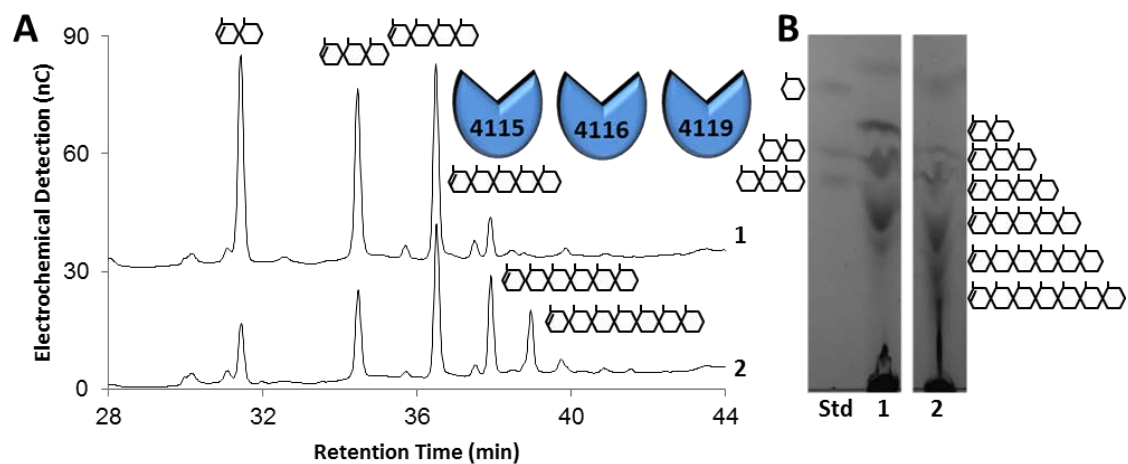


Figure 2.12: HPAEC-PAD and TLC analyses of PL1 matrix digestions. (A) HPAEC-PAD analysis of BT4115, BT4116 and BT4119 HG co-digestion. Traces 1 and 2, represent 24 and 1 h time points, respectively. **(B)** TLC analysis of BT4115, BT4116 and BT4119 HG digestion. Lanes 1 and 2, reaction time points (24 and 1 h, respectively). Std- GalA, GalA₂ and GalA₃ standards. Product peaks and standards are identified by corresponding carbohydrates in the margins.

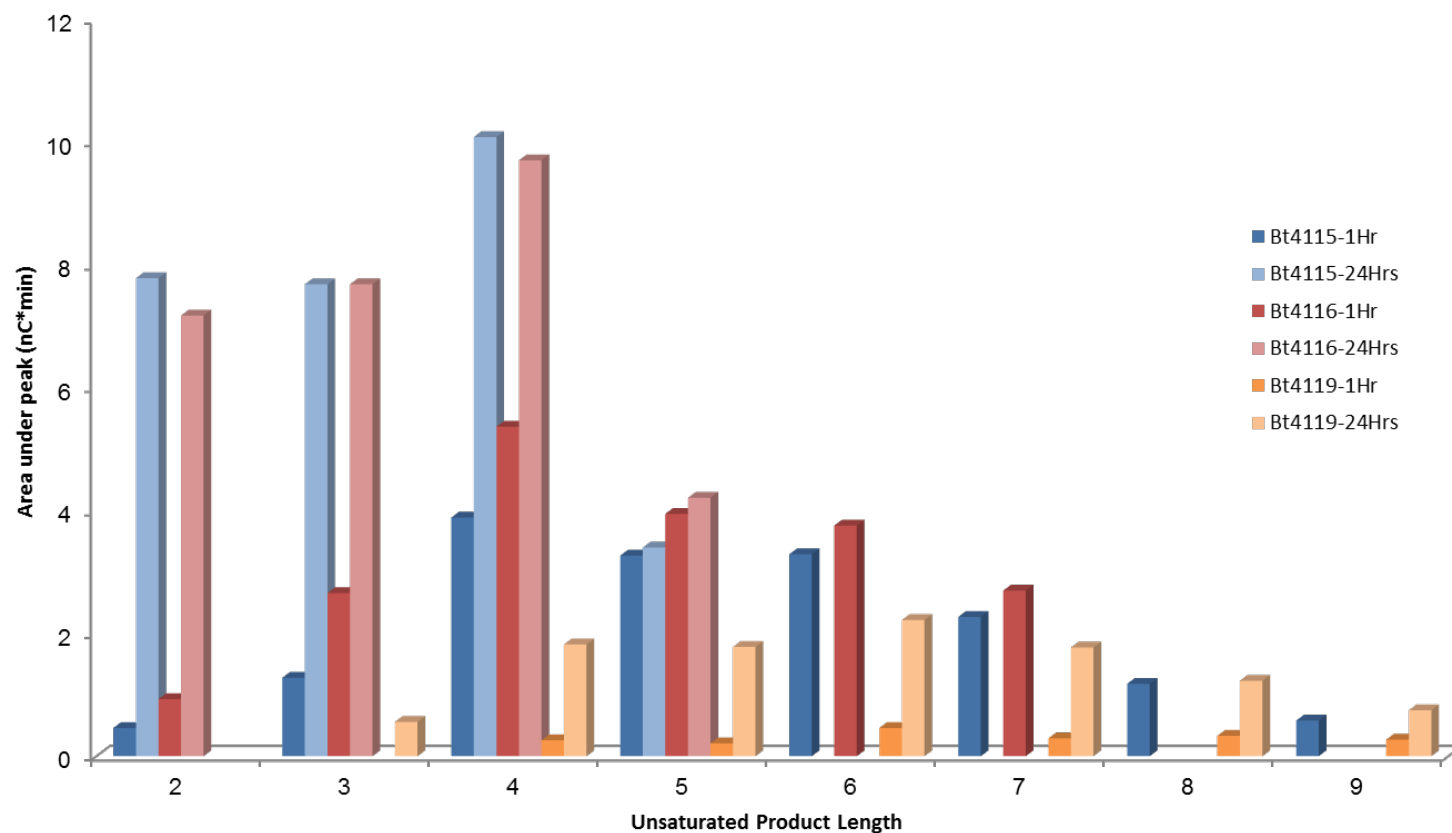


Figure 2.13: Graphical representation of area under individual unsaturated product peaks from HPAEC-PAD analysis of BT4115, Bt4116 and BT4119 HG time course digestions. Area under the peaks was calculated using the Chromeleon™ software. Enzymes present in the digestion are indicated in the legend. On the x-axis the length of unsaturated oligogalacturonides are indicated and on the y-axis the area under the product peak is indicated.

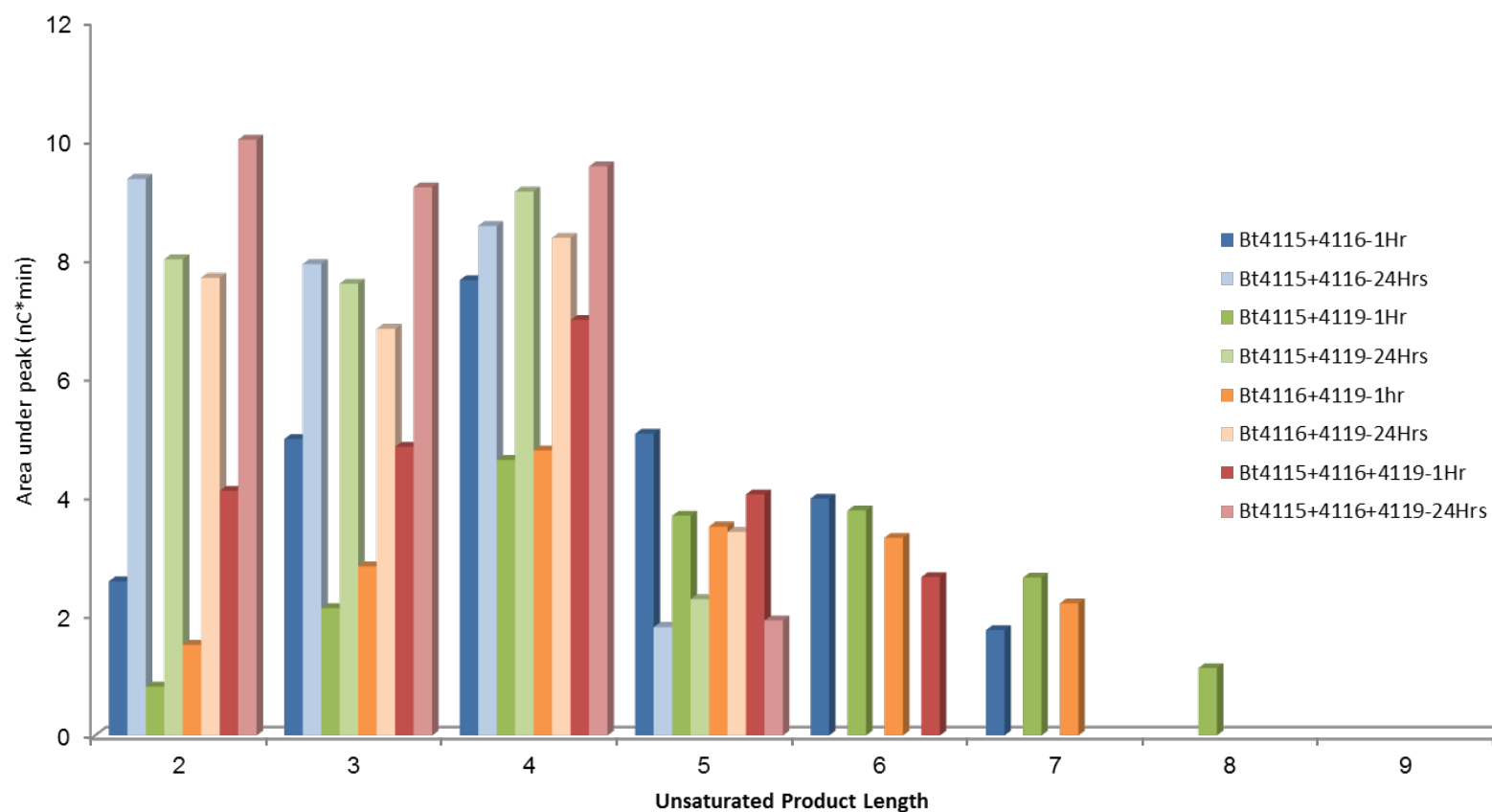


Figure 2.14: Graphical representation of area under individual unsaturated product peaks from HPAEC-PAD analysis of BT4115, Bt4116 and BT4119 HG time course digestions. Area under the peaks was calculated using the Chromeleon™ software. Enzymes present in the digestion are indicated in the legend. On the x-axis the length of unsaturated oligogalacturonides are indicated and on the y-axis the area under the product peak is indicated.

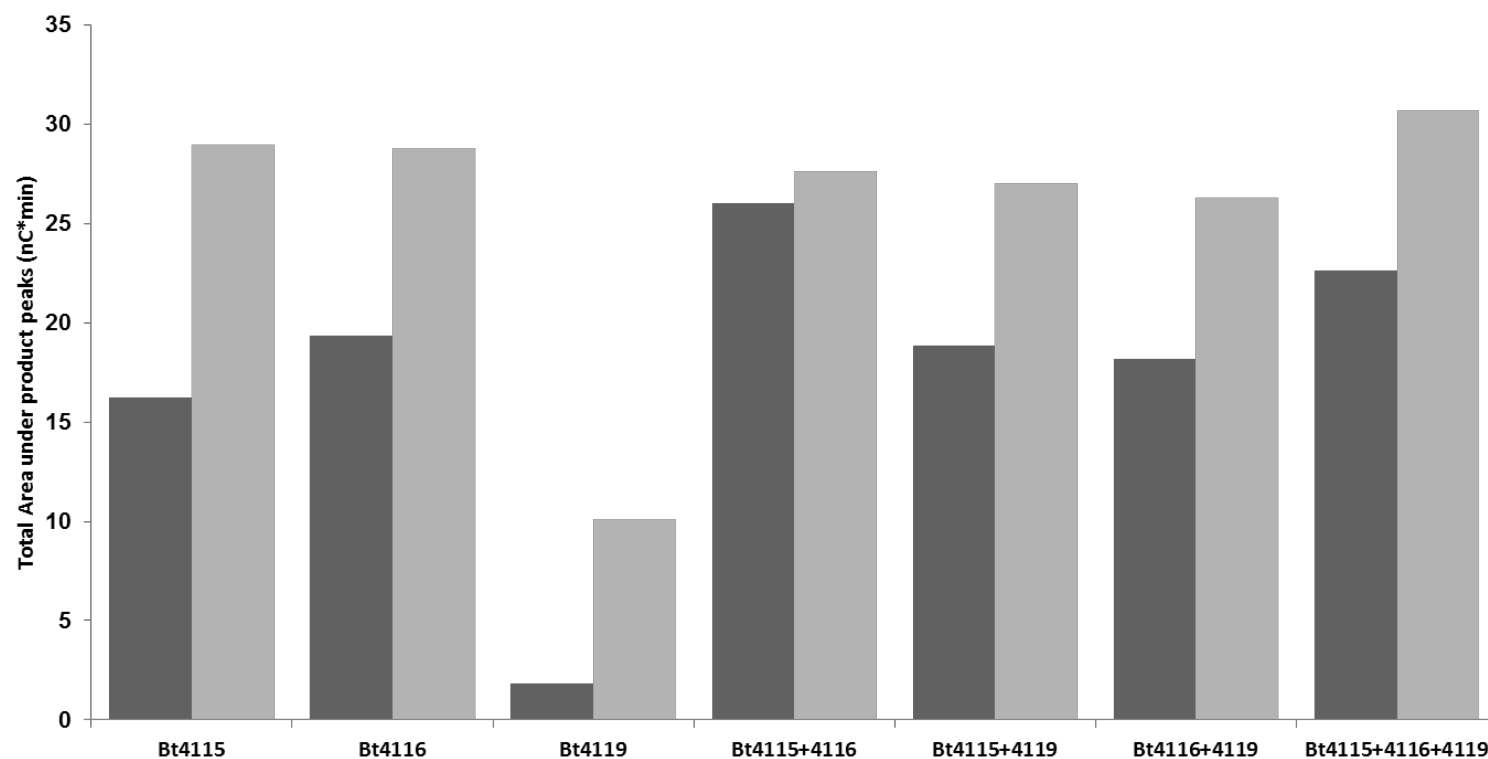


Figure 2.15: Graphical representation of total area under product peaks from HPAEC-PAD analysis of BT4115, Bt4116 and BT4119 HG time course digestions. Area under the peaks was calculated using the Chromeleon™ software. Enzymes present in the digestion are indicated on the x-axis with total area under the product peaks indicated on the y-axis. Dark grey and light grey represent 1 and 24 h time points, respectively.

2.4 Discussion

Though it is not uncommon for multiple copies of enzymes from the same family to be present in the genome of an organism, the presence of multiple copies of familial enzymes within one PUL is interesting. As such, the presence of three PL1s in PUL75 was unclear. Do these enzymes display unique activities and product profiles, or do they represent redundant copies of a common enzyme? In order to elucidate the individual roles of the three PL1s present in PUL75, we sought to clone and to produce recombinant protein to facilitate the characterization of these putative enzymes. In this light we cloned the coding sequences from *BT4115*, *BT4116* and *BT4119* into pET28a or pET32a vector for recombinant protein expression (Figure 2.2A-C). The recombinant proteins were purified by IMAC and eluted at high concentrations of imidazole (Figure 2.2D-F).

The pH optimum of BT4115 was determined using a buffer series spanning a pH range of 6.8-10.4. Activity dropped off sharply below 7.4 and above 9.4, with optimal activity within that range (7.5-9.5; Figure 2.3). This pH range is consistent with previously characterized PL families (1, 3 and 9)⁹⁰. The alkaline activity of pectate lyases has been attributed to a catalytic arginine and its participation in proton abstraction and localized pKa environments⁹⁹.

A divalent cation, typically Ca^{2+} , is a requirement for PL1 activity. As such we sought to investigate BT4115 activity in the presence of Ca^{2+} and other divalent cations (Mg^{2+} , Mn^{2+} , Co^{2+} , Ni^{2+} and Zn^{2+}) (Table 2.2). Consistent with previous PL1 characterization BT4115 displayed a preference for Ca^{2+} displaying a k_{cat} and K_{m} of $3800 \pm 280 \text{ min}^{-1}$ and $0.19 \pm 0.04 \text{ mg ml}^{-1}$, respectively. The resulting effects of cations on enzyme activity can

be described in this manner; $\text{Ca}^{2+} > \text{Mg}^{2+} > \text{Mn}^{2+} > \text{Co}^{2+} > \text{Ni}^{2+}$. The kinetic analysis with Zn^{2+} was not obtained due to complete precipitation of BT4115 during dialysis. Interestingly, a subsequent sample that was not dialyzed with Ca^{2+} , but was supplemented with 0.1 mM Ca^{2+} during kinetic analysis, displayed a k_{cat} and K_{m} of $1200 \pm 110 \text{ min}^{-1}$ and $0.033 \pm 0.013 \text{ mg ml}^{-1}$, respectively. At first glance, it appears the dialyzed sample to be more enzymatically active, however, when comparing the catalytic efficiency ($k_{\text{cat}}/K_{\text{m}}$), we see an approximately 2-fold difference between the two samples $20000 \pm 5700 \text{ min}^{-1}/\text{mg ml}^{-1}$ compared to $37000 \pm 18000 \text{ min}^{-1}/\text{mg ml}^{-1}$. A brief explanation for this observation can be attributed to the calcium coordination abilities of the HG substrate. This effect is commonly seen in PL1 kinetic analyses, and could be described as metal sequestration⁹⁰. As the substrate concentration increases, so do the number of calcium coordination sites. The increased number of calcium coordination sites between HG substrate molecules provides less calcium for enzymatic activity. Interestingly, when calcium was dialyzed in the presence of enzyme, before substrate addition, this metal sequestration effect was not observed (Figure 2.5A).

A time course digestion with these different divalent cations was performed and the products analyzed by TLC (Figure 2.5B). The product profiles observed after 1 and 24h digestion are consistent with the observed kinetic data. Namely, the greater the observed rate (Figure 2.5A), the shorter the length of products observed in the TLC analysis (Figure 2.5B). The observed differences in enzymatic activities between the divalent cations examined can be explained by analyzing the ionic radii. The size of the ionic radii exhibits the same trend as the observed activities, $\text{Ca}^{2+} > \text{Mg}^{2+} > \text{Mn}^{2+} > \text{Co}^{2+} >$

Ni^{2+} , suggesting that the size of the cation is directly related to the observed activities and corresponding product profiles.

Subsequent digestions of HG with the three different PL1s (BT4115, BT4116 and BT4119) confirmed endolytic activity, as at different time points in the reaction coordinates unsaturated oligogalacturonides of varying lengths (uGalA_n , where n is degree of oligomerization) are produced (Figure 2.7). Interestingly, the product profiles of the individual HG digestions provided the first insights into the possibility of distinct roles for the three PL1s within PUL75. At completion, two of the three PLs (BT4115 and BT4116) displayed product profiles consisting of three main unsaturated oligogalacturonides, uGalA_2 , uGalA_3 and uGalA_4 (Figure 2.7A & B). The third PL1, BT4119, displayed production of a greater proportion of uGalA_5 and uGalA_6 products as compared to the others. At completion, all three product pools lack either saturated (GalA) or unsaturated (5-keto-4-deoxyuronate; DKI) monosaccharides. These observations provide insight into the number of subsites within the active site of each enzyme (Figure 2.7C). More specifically, the BT4115 active site contains a minimum of 5 subsites, as uGalA_4 is present after exhaustive digestion. Similarly, BT4119 reveals a minimum of 7 subsites, as uGalA_6 is present after exhaustive digestion. Based on its product profile, we believe the two most plausible subsite orientations result in the production of uGalA_3 and uGalA_4 products. At first glance, BT4116 would appear to have the same number of subsites as BT4115, due to the similar product profile, however upon further analysis of BT4116 digestion with BT4115 HG digestion products (uGalA_2 , uGalA_3 and uGalA_4) (Figure 2.9A), we observed only uGalA_2 and uGalA_3 present. This indicates a minimum of 4 subsites in the BT4116 active site, and the corresponding

model of subsite orientation (Figure 2.9C). Interestingly, HPAEC-PAD analysis of BT4119 products did detect the presence of smaller unsaturated oligogalacturonides (Figure 2.10C); however, the quantification of these products was approximately 2- and 3-fold less when compared with the other PL1s, whereas the larger oligogalacturonides, uGalA₅ and uGalA₆ were approximately 2-fold more. These observations suggest an inherent order between the three PL1s in the deconstruction of HG, namely, BT4119 is further upstream in the process than either BT4115 or BT4116, and BT4115 and BT4116 activities on HG result in the generation of short unsaturated oligogalacturonides. To investigate further, we performed individual time course analyses of all three PL1s (Figure 2.7). Consistent with our previous results, we observed that BT4119 preferentially produced longer unsaturated oligogalacturonides as compared to BT4115 and BT4116. This preference for the production of longer unsaturated oligogalacturonides is consistent with our placement of BT4119 ‘upstream’ of either BT4115 or BT4116 in HG deconstruction. In the early time points, BT4115 produces longer unsaturated oligogalacturonides, uGalA₅ and uGalA₆, however these products are subsequently digested into the shorter unsaturated oligogalacturonides. Unlike either the BT4115 or BT4119 time courses, the BT4116 time course revealed the production of little to no products during the initial time points. Not until the 2-3 h mark do we see the generation of uGalA₃ (Figure 2.7C). This would suggest that BT4116 may prefer shorter length substrates, as compared to BT4115 and BT4119.

To investigate complementary activities between these enzymes, explaining their combined presence in PUL75, we performed digestions with the products produced from individual PL1 HG digestions (Figure 2.9A & B). Analysis of these overnight digestions

by TLC and HPAEC-PAD confirmed complementary activities between BT4119 and BT4115, and BT4115 and BT4116. An increase in the shorter unsaturated oligogalacturonide fractions upon subsequent digestions is difficult to distinguish in the TLC analysis; however, the HPAEC-PAD analysis provides quantification of this increase in shorter oligogalacturonide production, with a corresponding decrease in longer oligogalacturonides (Figure 2.9B). To further quantify the observed changes in product profiles with addition of complementary enzymes, we set up a digestion matrix as indicated in Table 2.1. This digestion matrix provided further insight into the complementary nature of the PL1s and the resulting product profiles by analyzing the 1 and 24 hr time points for the various combinations of PL1s via TLC and HPAEC-PAD (Figures 2.10-12). It should be noted that the rate at which these products were produced was not included as an observable in these experiments, therefore, the matrix digestions are not to be interpreted as a kinetic analysis of complementary PL1 activities. In all of the analyzed digestions, with an increase in time, we see a distinct shift to the production of smaller unsaturated oligogalacturonides. In an effort to quantify these products, we analyzed the area under the peaks for each of the chromatograms. Not surprisingly, in the digestion with all three PL1s present, we observed the greatest combined area under the peaks (Figure 2.15). In the product profiles of the individual PL1 HG digestions we observed the smallest combined area under the peaks, indicating the least amount of products produced (Figure 2.13). Generally speaking, when two of the three PL1s were present, the amount of products produced was greater than in the individual digestions, however it did not exceed that with all three PL1s present (Figures 2.13-15). The trends from our previous complementary digestions were further supported, but most

importantly, the amount of uGalA₂ produced in the BT4115 and BT4116 co-digestion was greater than all other co-digestions, save for in the presence of all three. This result confirmed our observation that BT4116 prefers shorter unsaturated oligogalacturonides, as polymerized HG appears to have an inhibitory effect.

In an attempt provide further insight into the spatial location of these enzymes, a number of predictive bioinformatic tools were employed. Most interestingly, BT4119 contained a predicted N-terminal transmembrane domain (Figure 2.16). This predicted domain is consistent with BT4119 being present as an outer membrane bound protein. Based on the observed product profile and by analogy to the SusG-like outer membrane depolymerase from the well characterized Sus locus in *B. thetaiotaomicron*, we believe that BT4119 is an outer membrane bound enzyme that depolymerizes HG. The resulting unsaturated oligogalacturonides are subsequently transported into the periplasmic space by the Sus-like transport machinery (see Chapter 4 for further discussion), for further deconstruction within the cell by the remaining periplasmic PL1s, BT4115 and BT4116. In this light, BT4119 acts as a ‘gate-keeper’ modulating that amount of unsaturated oligogalacturonides available for further deconstruction and is an integral component of PUL75. In order to test this hypothesis future experiments will knockout the *BT4119* gene, to determine the effect on *B. thetaiotaomicron* grown on polymerized HG. If our hypothesis is correct, the mutant should display significantly impaired growth or no growth at all.

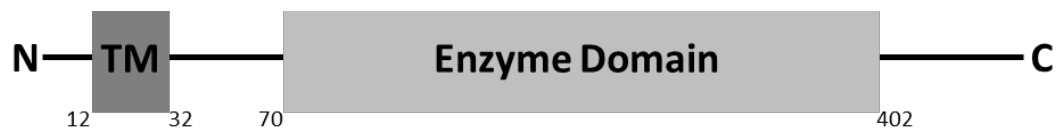


Figure 2.16: Cartoon representation of BT4119 predicted domains. The full length BT4119 protein sequence was inputted into the InterproScan server, resulting in the following output. N and C indicate the N- and C- termini of the protein sequence, and the numerical values indicate domain boundaries. TM-Transmembrane domain.

2.5 Conclusion

The successful cloning, recombinant protein production, and enzymatic characterization of the three PL1s present in PUL75, namely BT4115, BT4116 and BT4119, confirmed all three gene products to be endolytic enzymes active on HG. Enzymatic activity was found to be Ca^{2+} dependent for all three PL1s. Furthermore, the successful characterization of the three PL1s provided evidence of complementarity between all three enzymes, providing a biological rationale for the multiple copies of PL1s present within PUL75. Based on our characterization, we propose that BT4119 is an outer membrane anchored endolytic enzyme responsible for the production of longer unsaturated oligosaccharides. These products are transported into the periplasmic space via the Sus-like transport machinery (see Chapter 4). Once transported, these unsaturated oligogalacturonides are further deconstructed by BT4115, producing shorter unsaturated oligogalacturonides that are the preferred substrate for BT4116 activity. A minimum of 4, 5 and 7 subsites for BT4116, BT4115 and BT4119 activity, respectively, were proposed based on the largest product of the individual product profiles. How these unsaturated oligogalacturonides are further processed by *B. thetaiotaomicron* will be determined in Chapter 3.

Chapter 3-

Elucidating the roles of Glycoside Hydrolases BT4108 and BT4123 in PUL75 mediated HG deconstruction

3.1 Introduction

Glycoside Hydrolases (GHs) comprise the largest group of CAZymes identified to date and are key to carbohydrate metabolism across all domains of life. As previously described, GHs are responsible for the hydrolysis of glycosidic linkages in glycosides. The sheer number of GHs characterized to date reflects the enormous diversity and complexity found among their target carbohydrate substrates. Already there are more than 130 different related families in the CAZy database, based on sequence homology. Among these, two have emerged as the primary agents of pectic polysaccharide (HG and RG-I) hydrolysis, GH28 and GH105. In both cases, these families have been shown to cleave linkages with different chemistries.

3.1.1 GH mechanism— GH28s operate by a conserved single-displacement mechanism resulting in inversion of stereochemistry of the anomeric carbon from C1- α to C1- β (Figure 3.1A). The first detailed description of the GH28 mechanism was from endopolygalaturonases I and II (*Aspergillus niger* and *Aspergillus tubingensis*) using reduced substrates¹⁰⁰. The reaction is catalyzed by a triad of aspartates clustered on the same side of the active site cleft (sometimes referred to as the

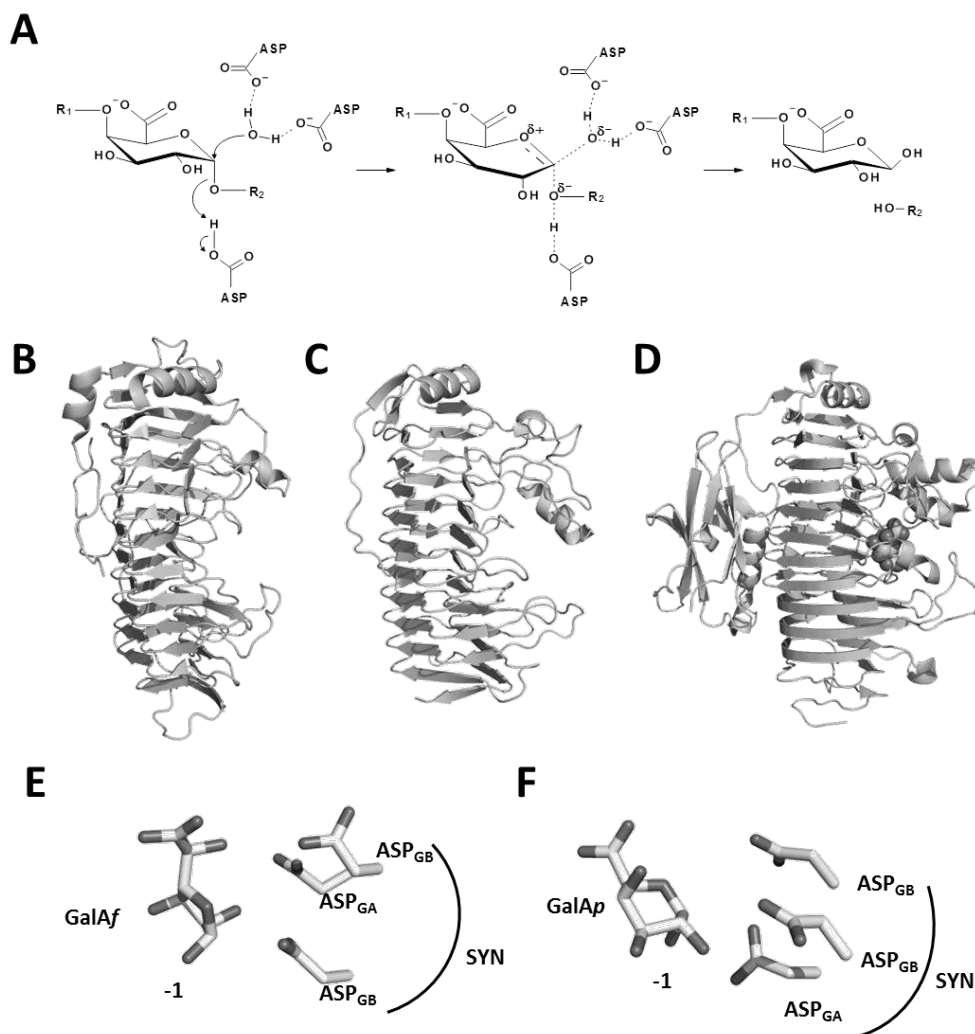


Figure 3.1: Structure and function of GH28s on pectic carbohydrates. (A) Inverting hydrolysis mechanism performed by GH28s. Three dimensional structures of (B) *A. aculeatus* KSM 510 / CBS 115.80 endorhamnogalacturonase (PDB ID: 1RMG)¹⁰¹, (C) *P. carotovorum* SCC3193 endopolygalacturonase (PDB ID: 1BHE)¹⁰², and (D) *Y. enterocolitica* spp. *enterocolitica* 8081 exopolygalacturonase (PDB ID: 2UVF)¹⁰³. Comparison of the three catalytic aspartates (i.e. ‘syn’ conformation) in GH28s from the -1 subsite of endopolygalacturonase I from *Stereum purpureum* with α -D-galacturofuranose (PDB ID: 1KCD)¹⁰⁴ (E) and exopolygalacturonase from *Y. enterocolitica* spp. *enterocolitica* 8081 with α -D-galacturonopyranose (PDB ID: 2UVF)¹⁰³ (F).

‘syn’ conformation) (Figure 3.1F). This architecture differs from the canonical tandem general acid and general base orientation observed in other inverting GHs, in which the residues are about 10 Å apart and opposed on either side of the substrate¹⁰⁵. Asp201¹⁰⁰ (Asp223 in *Pectobacterium carotovorum* PehA¹⁰² / Asp402 in exoGH28 in *Y. enterocolitica*¹⁰³) is believed to function as the general acid by donating a proton to the glycosidic oxygen. The other two aspartates, 180 (202/381) and 202 (224/403), which have been identified as general bases, interacting with the nucleophilic water and charging it for attack of the anomeric carbon.

GH105s use a unique mechanism that differs from the canonical inverting and retaining mechanisms of most GHs (Figure 3.2A). The structure of YteR from *Bacillus subtilis* strain 168 (Figure 3.2B), an unsaturated rhamnogalacturonyl hydrolase, in complex with unsaturated uGlcA-(α -1,4)-GalNAc¹⁰⁶ (u - unsaturated) and an uGalA-(α -1,2)-Rha¹⁰⁷ (uGlcA and uGalA are sterically identical) identified Asp143 as the general acid, and His189 as a general base that activates a catalytic water (Figure 3.2C). Significantly, the scissile bond in the hydrolytic reaction is the C4 and C5 alkene as opposed to the glycosidic bond. The reaction progresses through hydration of the vinyl ether. Asp143 donates a proton to the C4 atom and the activated water attacks the C5 creating an unstable hemiacetal. This compound decomposes into the linear 5-keto-4-deoxyuronate (DKI) and releases the glycone leaving group.

3.1.2 GH Activity- GH28s and GH105s contain diverse activities for enzymes that are active on functionally related but chemically distinct pectins (e.g. α -1,4 galacturonosyl and α -1,2 rhamnosyl containing substrates). Characterized members from GH28 include polygalacturonase (EC 3.2.1.15)¹⁰⁸, exopolygalacturonase (EC 3.2.1.67)¹⁰⁹,

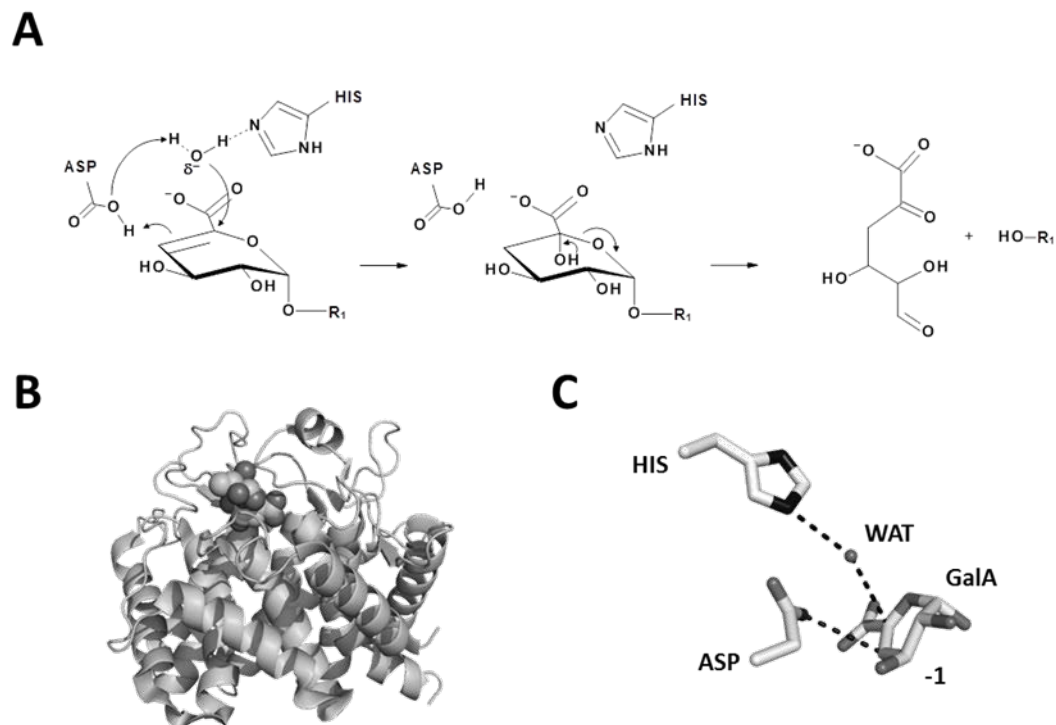


Figure 3.2: Structure and function of GH105s on unsaturated pectic carbohydrates. **(A)** Vinyl ether hydrolysis mechanism catalyzed by GH105s. The catalytic Asp and His involved in hydration of the unsaturated C4-C5 bond in GalA are shown. The linearized product, DKI, is seen on the right. **(B)** Cartoon representation of the *Bacillus subtilis* subsp. *subtilis* str. 168 in complex with α -D-4-deoxy-GlcA-(1-2)- α -L-Rhap (PDB ID: 2GH4)¹¹⁰. **(C)** Catalytic residues within the -1 subsite involved in vinyl ether hydrolysis.

exopolygalacturonosidase (EC 3.2.1.82)¹⁰⁹, rhamnogalacturonase (EC 3.2.1.171)¹¹¹, rhamnogalacturonan α -1,2-galacturonohydrolase (EC 3.2.1.173)¹¹², and rhamnogalacturonan α -L-rhamnopyranohydrolase (EC 3.2.1.174)¹¹³. These coordinated activities have the capacity to saccharify complex polymerized HG and RG-I into GalA and Rha monosaccharides, and many of these activities have been harnessed for food processing applications^{114, 115}. GH105s are exolytic enzymes that remove terminal uGalA from the products of PLs (EC 3.2.1.172). The first activity was observed in a *Bacillus subtilis* subsp. *subtilis* str. 168 enzyme that was specifically active on RG lyase products^{106, 107}. More recently a GH105 homolog has been described from the green macroalgae *Nonlabens ulvanivorans* (NdGH105)¹¹⁶. This enzyme harnesses a similar vinyl-ether hydration mechanism to cleave a unique β -uGlcA-(1-4)-Rha-(sulphate)₃ linkage, indicating that this mechanism is not exclusive to a defined stereochemistry of the anomeric carbon. NdGH105 is active on the β -uGlcA moiety present within the algal cell wall polysaccharide ulvan (3-sulfated rhamnose, glucuronic acid, iduronic acid, and small amounts of xylose) that had been treated with ulvan lyases. The determination that GH105s are active on both α - and β -configured glycosidic linkages highlights the plasticity within their active sites, and suggests the possibility that a wide spectrum of activities on lyase products may yet be discovered within this family.

3.1.3 GH structural highlights— GH28 and GH105 adopt distinct folds (Figures 3.1 & 3.2). GH28s have right-handed parallel β -helices that differ from the β -helix of PLs by displaying a four-sided β -sheet architecture as opposed to three. The first family structure solved was the rhamnogalacturonase from *Aspergillus aculeatus* (Figure 3.1B)¹⁰¹, which

was followed by a bacterial endopolygalacturonase from *Pectobacterium carotovorum* SCC3193 (Figure 3.1C)¹⁰². Structural superimposition of these two enzymes revealed that they are very similar in overall structure; however, the bacterial endopolygalacturonase had a shortened topology with one fewer β -helix turn and a unique C-terminal cap¹⁰². Elucidating the structural basis of exopolygalacturonase activity took nearly a decade after these seminal insights. YeGH28, which is a disaccharide releasing exopolygalacturonosidase (EC 3.2.1.82), provides an example of large-scale changes to the active site cleft transforming endolytic to exolytic activity (Figure 3.1D). Four loop insertions converge to form a blind canyon, which restricts access of substrate to one orientation and results in exclusive production of disaccharide products¹⁰³. The structure of a monosaccharide releasing exopolygalacturonase (EC 3.2.1.67) from *Thermotoga maritima* revealed that GH28s can oligomerize to increase stability under thermophilic conditions¹¹⁷. Product complexes for an endopolygalacturonase (-1 and +1 subsites)¹⁰⁴ and an exopolygalacturonase (-1 and -2 subsites) have helped to define the subsite architecture within GH28s (Figure 3.1E-F)¹⁰³.

Several structures have been deposited for bacterial homologs of GH105. YteR exhibits an α/α double-toroid structure with six α -hairpins arranged in a double α -helical barrel (Figure 3.2B)¹¹⁸. Overlays of YteR complexes with uGlcA-(1-4)-GalNAc¹⁰⁶ and GalA-(1-2)-Rha¹⁰⁷ revealed that the anhydro subunit is positioned in nearly identical orientations. Superimpositions with a GH88 complex, which are related β -unsaturated glucuronyl hydrolases active on the products of chondroitin lyases, confirmed that the catalytic machinery is conserved between these homologs, and suggests a distant relatedness and conserved mechanism¹⁰⁶.

PUL75 from *B. thetaiotaomicron* contains a putative enzyme from each of these two families; BT4108 (GH105) and BT4123 (GH28), however, the activities and substrate specificities of these two gene products have yet to be determined. In this light, these genes were targeted for cloning, expression, and recombinant protein purification in order to characterize their activities and elucidate their roles within PUL75 of *B. thetaiotaomicron*.

3.2 Materials and Methods

3.2.1 Recombinant protein production— The coding sequences of the two putative glycoside hydrolase families from PUL75, *BT4108*-GH105 and *BT4123*-GH28, were amplified from *B. thetaiotaomicron* genomic DNA using engineered primers (Appendix 1). Amplicons were restricted with NheI (5') and XhoI (3') enzymes and ligated into pET28a (Novagen, Cat #69864-3) or NcoI (5') and XhoI (3') pET32a (Novagen, Cat #69015-3) expression vectors with complementary ends to create the expression plasmids pET28-BT4108 and pET32-BT4123. These constructs contain N-terminal 6x histidine cleavable tags for purification by immobilized metal affinity chromatography (IMAC). All expression plasmids were individually transformed into BL21 (DE3) (EMD Millipore, Cat #69450-3, Karmstadt, Germany) cells for protein overexpression. These transformants were used to inoculate 1 l of LB broth containing kanamycin (50 $\mu\text{g ml}^{-1}$: pET28a) or ampicillin (50 $\mu\text{g }\mu\text{l}^{-1}$: pET32a). Cultures were grown at 37°C to an OD₆₀₀ of 0.8-1.0, cooled to 16°C and induced with 0.2 mM IPTG (Cells were continuously shaken at 180 rpm during the growth and induction process). The following day the cells were harvested by centrifugation at 6,500 x *g* for 10 min at 4°C. The cell pellet was re-suspended in sucrose solution [50 mM tris-HCl (pH 8.0), 25% sucrose with 10 mg l⁻¹ lysozyme] and chemically lysed at ambient temperature with 2 volumes of lysis buffer [1% deoxycholate, 1% triton X-100, 20 mM tris-HCl (pH 7.0), and 100 mM NaCl] with 200 $\mu\text{g l}^{-1}$ DNase. The lysate was clarified by centrifugation at 17,500 x *g* for 45 min at 4°C and filtered using a 0.45 μM low protein binding syringe filter (Pall Acrodisc 25mm). This soluble protein fraction was applied to a nickel (II) nitrilotriacetic acid (Ni²⁺-NTA) sepharose support (GE Healthcare, Cat#: 17-5318-06) and the immobilized

protein was washed with 500 mM NaCl, 20 mM tris-HCl (pH 8.5) to decrease non-specific binding. A stepwise imidazole gradient (5 mM-500 mM) was used to elute the target protein. Samples were visualized by SDS-PAGE, and fractions with appreciable amounts of target protein were pooled and dialyzed overnight against 4 l of 20 mM tris (pH 8.0), with two buffer exchanges, using a dialysis membrane (Spectrum, Biotech CE) with 5,000 MW cutoff. The following day protein concentration was determined at OD_{280nm} using extinction coefficients of 94,575 M⁻¹ cm⁻¹ (2.057 mg ml⁻¹) and 66,850 M⁻¹ cm⁻¹ (1.151 mg ml⁻¹), respectively, as calculated by the ProtParam analysis tool ¹¹⁹.

3.2.2 GH105 pH optimum— To investigate GH105 pH optimum, 200 μM uGalA₂ was incubated with 1.86 μM enzyme in 20 mM of one of the following buffers; sodium acetate trihydrate (pH 4), sodium citrate tribasic dehydrate (pH 5), sodium cacodylate trihydrate (pH 6), hepes sodium (pH 7) and tris hydrochloride (pH 8). Enzyme activity was measured at 232 nm, detecting the loss of unsaturation as the unsaturated bond is saturated and the ring linearized¹⁰⁶.

3.2.3 GH105 kinetic analysis— GH105 kinetics were obtained by spectrophotometric analysis of the loss of unsaturation resulting from the digestion of enzyme with varying concentrations of uGalA₂ (Ext Co of 4000 M⁻¹ cm⁻¹) measured at 232 nm. Reaction volume totaled 600 μl with final concentrations of 20 mM sodium acetate (pH 4.2), 90 nM enzyme and substrate (as indicated). The linear portion of the reaction curve was used to determine the rate of hydrolysis. Data points were collected in triplicate, with the average calculated and processed with Graph Pad Prism 6™.

3.2.4 GH28 pH optimum— Time course digestions were performed with GH28 to determine pH optimum. Reactions were prepared as above with the addition of 20 mM buffer: sodium acetate trihydrate (pH 4), sodium citrate tribasic dehydrate (pH 5), sodium cacodylate trihydrate (pH 6), hepes sodium (pH 7) and tris hydrochloride (pH 8). Time points were taken at 5, 30, 60, 180 and overnight and heat treated at 100°C for 10 min. Samples were centrifuged and spotted onto TLC plates, products separated and visualized as above.

3.2.5 GH28 kinetic analysis— A variant of the 3,5-dinitrosalicylic acid (DNSA) reducing end assay¹²⁰ was employed for GH28 kinetic analysis. The DNSA reagent (1% NaOH, 1% DNSA, 0.2% Phenol, 0.002% Glucose, and 0.05% NaSO₃) was incubated in a 1:1 ratio with enzyme digestion samples. The solution was boiled for 5-10 min and placed on ice to cool to room temperature. Samples were plated in a 96-well plate and absorbances were obtained at 575 nm using a thermoregulated Biotek™ plate reader. Rates were calculated from the replicate trials at a given substrate concentration and Graph Pad Prism 6™ was used to process the data.

3.2.6 HG, unsaturated oligogalacturonide (uGalA_n) and saturated oligogalacturonide (GalA_n) digestions— Overnight digests of substrate were performed with each enzyme individually to determine specificity and to generate product profiles. BT4108 and BT4123 were incubated at 37°C with 1 mg ml⁻¹ substrate (HG, uGalA_n or GalA_n) in 20 mM sodium acetate pH 4.2 or sodium citrate tribasic dihydrate pH 5, respectively. BT4108 was also incubated overnight with a uGalA₂ sample obtained from the overnight digestion of GalA₃ with a PL2 from *Vibrio vulnificus*. This substrate contains uGalA₂

and GalA, and was prepared by Richard McLean (Ph.D. candidate in Abbott Lab). After incubation all samples were heat treated at 100°C for 10 min to denature the enzyme and terminate the reaction. The samples were centrifuged to remove precipitate and spotted onto a silica plate with aluminum backing for thin layer chromatography (TLC; EMD, Cat #55553-7). GalA (Sigma Cat# 48280), GalA₂ (Sigma, Cat# D4288) and GalA₃ (Sigma, Cat# T7407) were used as standards. The plates were run twice in a mobilization buffer consisting of 1-butanol, distilled water and acetic acid (5:3:2) and stained with a solution of ethanol and sulfuric acid (70:3 v/v) with 4% orcinol. The plate was then incubated at 120°C for 2-5 min in a dry oven for visualization.

3.2.7 HPAEC-PAD product analysis— HPLC runs were performed using a CarboPac™ PA-20 column (DIONEX, Cat #060142, Sunnyvale, CA) and a 0 – 900 mM NaOAc gradient in 100 mM NaOH with a flow rate of 0.5 ml min⁻¹. Samples were boiled, centrifuged, frozen prior to analysis, and diluted 1:5-1:20 (250 µL final volume) before loading. Retention times were determined and compared to GalA, GalA₂ and GalA₃ standards.

3.2.8 Sodium borohydride (NaBH₄) treatment of GalA₃ substrate— The reducing end of GalA₃ was selectively reduced with NaBH₄ in a two part reaction, to investigate the role of the reducing end in enzyme-substrate recognition. GalA₃ (2.7 mg) was re-suspended in acid and treated with 40 mg NaBH₄. The reaction was left on ice for 3 h for complete reduction of the GalA₃ reducing end. A control sample was solubilized in acid, but not reduced, and incubated on ice for 3 h. Both the reduced and control sample followed the same work-up, slow addition of glacial acetic acid, and were washed repeatedly with

ethanol. Following the washes, the samples were lyophilized, and re-suspended in equal volumes of distilled water. These substrates were used in overnight BT4123 digestions (as above).

3.2.9 GH105, PL1 and GH28 matrix digestions— A digestion matrix with BT4108, BT4115 and BT4123 was set up according to Table 3.1. Digestions conditions included 20 mM tris (pH 8) or 20 mM sodium citrate tribasic dehydrate (pH 5), 0.1 mM CaCl₂, 1 mg ml⁻¹ HG and 1 μM enzyme. A third reaction began at pH 8, however after overnight incubation, concentrated HCl was added to reduce the pH to 5. Reactions were incubated at 37°C with aliquots taken at 5, 15, 30, 60, 180, 300 min and 1-day. These aliquots were terminated by treatment at 100 °C for 10 min and analyzed via TLC.

Table 3.1: Matrix digestion template to investigate HG deconstruction by the different enzymes within PUL75

Reaction		i	ii	iii	iv	v	vi	vii
Enzyme	Class	<i>Enzymes Added</i>						
BT4108	GH105	X	X		X	X		
BT4115	PL1	X	X	X			X	
BT4123	GH28	X		X	X			X

3.3 Results

3.3.1 BT4108

3.3.1.1 Cloning, protein expression and purification— The coding sequence of BT4108 without the signal peptide was cloned into the pET28a *E. coli* expression vector in order to produce recombinant protein for characterization. Gene sequencing confirmed no mutations in the construct. BT4108 was expressed as an approximately 46 kDa soluble protein and purified by Immobilized Metal Affinity Chromatography (IMAC). BT4108 was pooled based on SDS-PAGE analysis (Figure 3.3C) and dialyzed. Approximately 10 mg of BT4108 was obtained from 1 L culture.

3.3.1.2 pH optimum— The pH optimum of BT4108 was determined by observing the decreased absorbance at 232 nm, resulting from the loss of the unsaturation at the non-reducing end. Initial pH optimum investigation included a range from 4-8; with no activity observed above pH 5. The pH optimum was further tested in the range of 3.6-4.6 (20 mM sodium acetate) with 4.0-4.2 displaying maximal activity (Figure 3.4A). A significant decrease in activity was seen above pH 4.2. This pH optimum is consistent with a previously characterized GH105¹⁰⁶.

3.3.1.3 Kinetic analysis— In an attempt to acquire kinetic parameters for BT4108, the hydration of the alkene in the unsaturated digalacturonic acid (uGalA₂) substrate was monitored at 232 nm at pH 4.2 with varying concentrations of substrate (Figure 3.4B). The kinetic analysis revealed a k_{cat} and K_{m} of $1.69 \pm 0.07 \text{ s}^{-1}$ and $82 \pm 11 \text{ }\mu\text{M}$, respectively. The rate of BT4108 activity can be described as moderate, as it is approximately 10 times greater than the observed rate for YteR; however, it is

approximately 7X less than slowest rate observed for NdGH105 activity. Complete kinetic parameters of BT4108 can be found in Table 3.2.

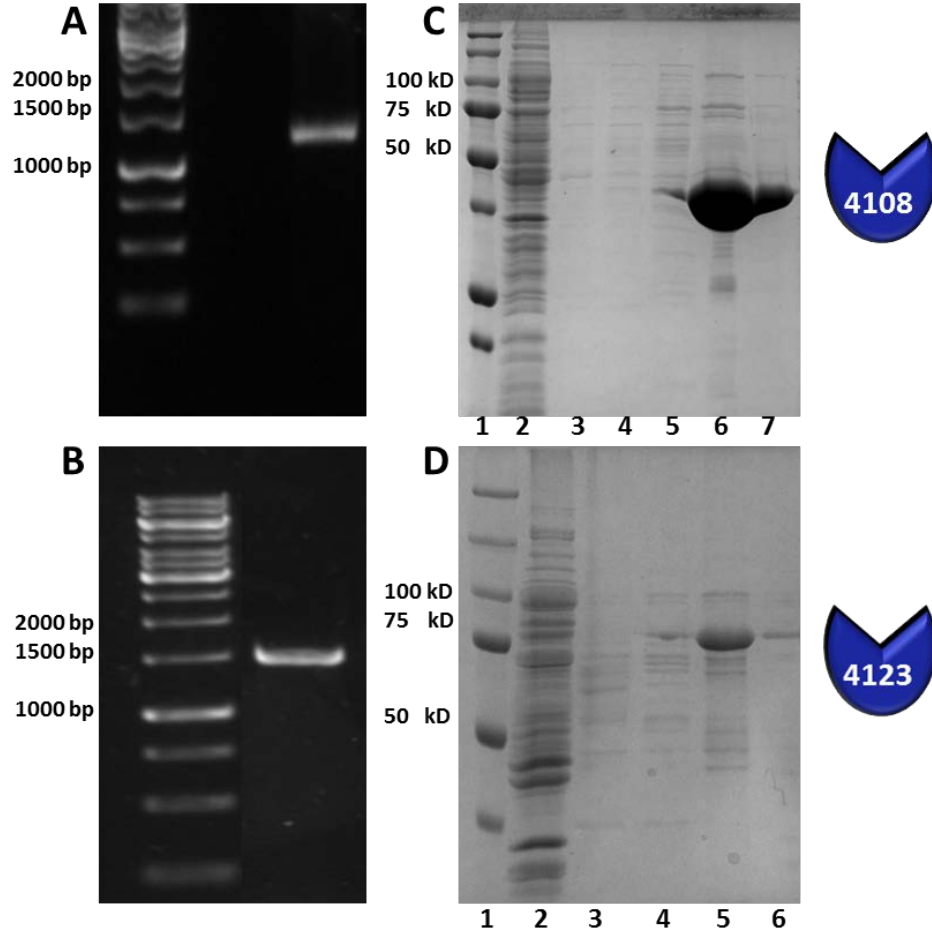


Figure 3.3: Cloning and recombinant protein purification of the two GHs from PUL75. (A-B). Agarose Gel (1%) displaying the PCR product for each of the GH genes, *BT4108* and *BT4123*, respectively. The MWM is on the left hand side of the gel, and the PCR product is on the right. **(C)** SDS-PAGE analysis of *BT4108* (pET28a) IMAC fractions. Lanes 1-7; MWM, cell lysate, wash, 5 mM, 10 mM, 100 mM and 500 mM, respectively. **(D)** SDS-PAGE analysis of *BT4123* (pET32a) IMAC fractions. Lanes 1-6; MWM, cell lysate, 5 mM, 10 mM, 100 mM and 500 mM elution, respectively.

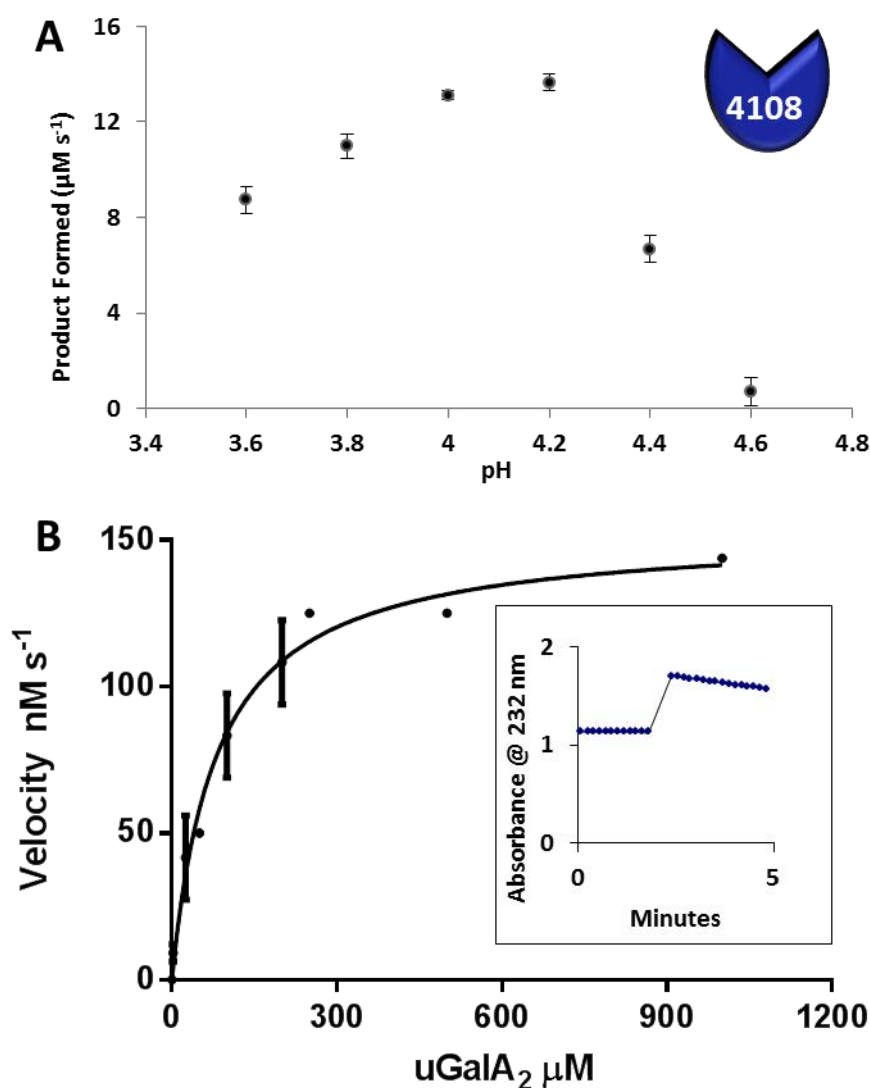


Figure 3.4: pH optimum and kinetic analysis of BT4108 (GH105) (A) pH optimum of BT4108 as measured by the loss of absorbance at 232 nm. Reactions were performed in triplicate with the average rate plotted. Reactions were run for 10 min and incubated at 37°C throughout the course of analysis. (B) Visual representation of the kinetic data obtained for BT4108 by the loss of absorbance at 232 nm. For a full description of kinetic parameters see Table 3.2. (Inset) Spectrophotometric trace of BT4108 digestion of uGalA_2

Table 3.2: Kinetic parameters obtained for BT4108 enzyme activity on uGalA_2

Enzyme	$V_{\max}$ (nM s^{-1})	K_m (μM)	k_{cat} (s^{-1})	k_{cat}/K_m ($\text{s}^{-1} \mu\text{M}^{-1}$)
BT4108 (GH105)	153 ± 7	82 ± 11	1.69 ± 0.07	0.021 ± 0.003

3.3.1.4 Product profile— Due to the large accumulation of uGalA_n products in PUL75, which are generated via PL1 activity, uGalA_n products generated from BT4116 digestion were incubated overnight with BT4108. The resulting products were analyzed by TLC and HPAEC-PAD. BT4108 displayed activity on uGalA_n (n=2 and 3) which was visualized by TLC (Figure 3.5A). A complete loss of the uGalA₂ is accompanied by the presence of a new GalA band (Figure 3.5A); however, the uGalA₃ migrates the same distance as GalA₂. In general, TLC analysis was unable to distinguish between GalA_n and uGalA_{n+1} products. As such, HPAEC-PAD was required to resolve and identify the resulting products. HPAEC-PAD analysis of the unsaturated oligogalacturonide (uGalA_n) products revealed a shift to corresponding saturated oligogalacturonides (GalA_n) upon treatment with BT4108 (Figure 3.5D). This indicates that BT4108 displays exolytic activity removing the unsaturated moiety from the non-reducing end. To further probe BT4108 activity, uGalA₂ substrate from PL2 digestion of GalA₃ (as described in 3.2.6), was incubated with BT4108. TLC analysis revealed the loss of the uGalA₂ band, and the subsequent increased intensity of the GalA band (Figure 3.5B). HPAEC-PAD analysis confirmed these initial TLC results, displaying the complete loss of the uGalA₂ peak (Figure 3.5E). These results confirmed that BT4108 displays exolytic activity, removing the unsaturated moiety from the non-reducing end, acting on a broad range of products down to uGalA₂.

3.3.1.5 uGalA_n time course— After initial activity screening, BT4108 was incubated with uGalA_n from terminated PL1-HG digestions and aliquots removed at given time points. Time point analysis revealed the generation of a GalA-sized product (Figure 3.6A); however, the larger GalA_n species remained undigested over the length of the time

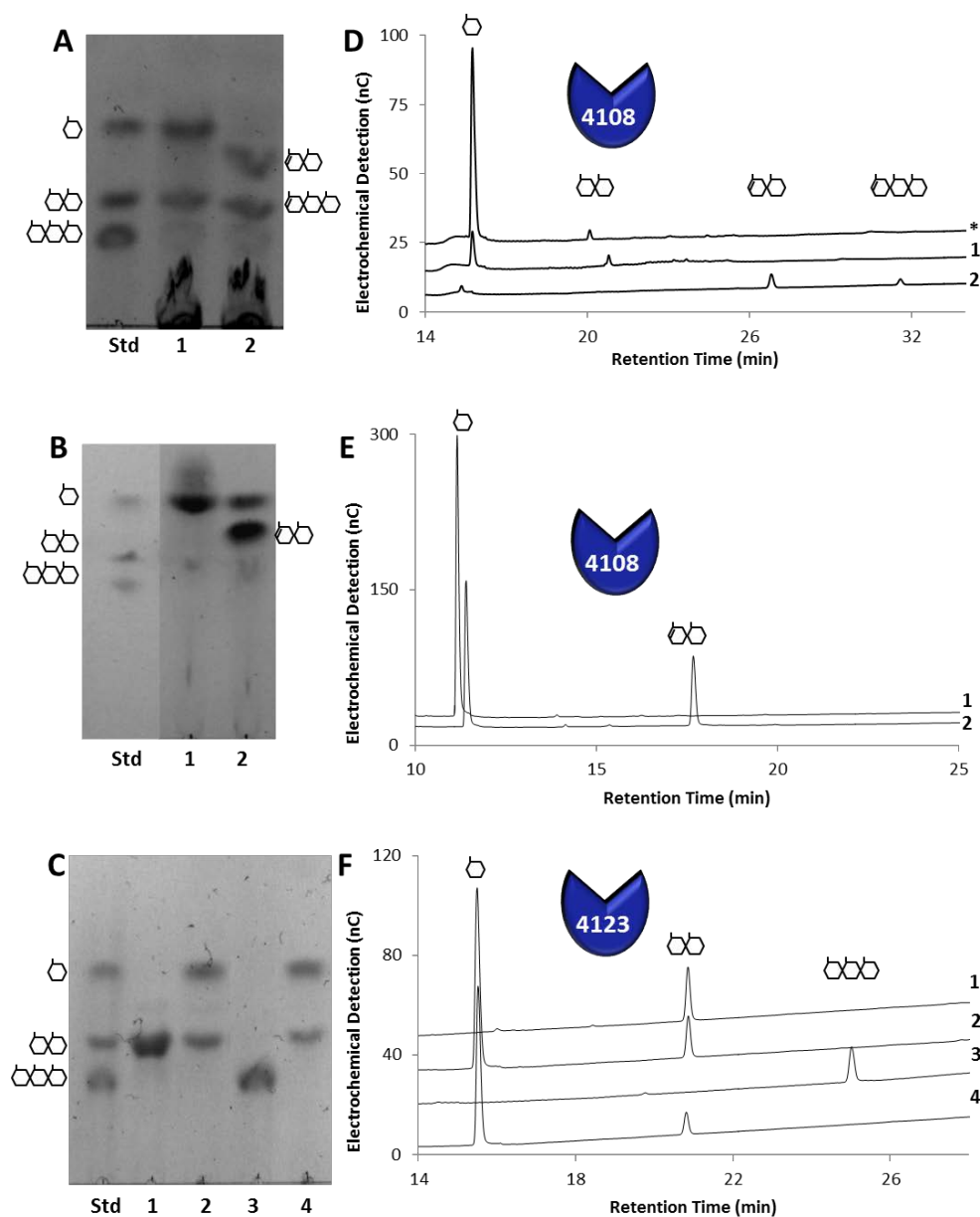


Figure 3.5: Profiles of oligogalacturonide digestion by GH105 (BT4108) and GH28 (BT4123). (A) TLC analysis of BT4108 (1 μ M) digestion of BT4116 unsaturated oligogalacturonide digestion products (1 mg ml⁻¹). Std- GalA, GalA₂, GalA₃ standards. Lane 1, overnight digestion products, lane 2 substrate only. Product sizes are indicated in the left (saturated) and right (unsaturated) margins. (B) TLC analysis of BT4108 (1 μ M) digestion of VvPL2 digestion products (1 mg ml⁻¹). Std- GalA, GalA₂, GalA₃ standards. Lane 1, overnight digestion products, lane 2 substrate only. Product sizes are indicated in the left (saturated) and right (unsaturated) margins. (C) TLC analysis of BT4123 (1 μ M) digestion of GalA₂ (lane 2) and GalA₃ (lane 4) (1 mg ml⁻¹). Lanes 1 and 3 are negative controls for GalA₂ and GalA₃, respectively.

Figure 3.5: cont.

All TLC samples were incubated overnight with and without enzyme. **(D)** HPAEC-PAD analysis of overnight BT4108 digestions as described in **A**. **(E)** HPAEC-PAD analysis of overnight BT4108 digestions as described in **B**. Trace indicated by a * corresponds to lane 1 of the TLC spiked with GalA (internal standard). **(F)** HPAEC-PAD analysis of overnight BT4123 digestions as described in **C**. All HPAEC-PAD trace numbers refer to corresponding samples from TLC lanes. Chromatogram traces have a number corresponding to the TLC lane of the analyzed sample. Product peaks are identified by corresponding carbohydrate.

3.3.2 *BT4123*

3.3.2.1 Cloning, protein expression and purification— The coding sequence of BT4123 was cloned into both the pET28a and pET32a *E. coli* expression vector in order to produce recombinant protein for characterization. Initial recombinant protein production was attempted using the pET28-BT4123 construct by Erin Crawford; however, minimal protein was produced. Subsequent cloning of *BT4123* into the pET32a fusion vector was successful in producing appreciable yields of recombinant protein. Approximately 3 mg of BT4123 was obtained from 1 L culture. Gene sequencing confirmed no mutations in the construct. BT4123 was expressed as an approximately 75 kDa soluble thioredoxin fusion protein (using pET32a) and purified by IMAC. BT4123 was pooled based on SDS-PAGE analysis (Figure 3.3D) and dialyzed. The thioredoxin fusion was not cleaved, and all subsequent analysis occurred with the fusion protein, referred to as BT4123.

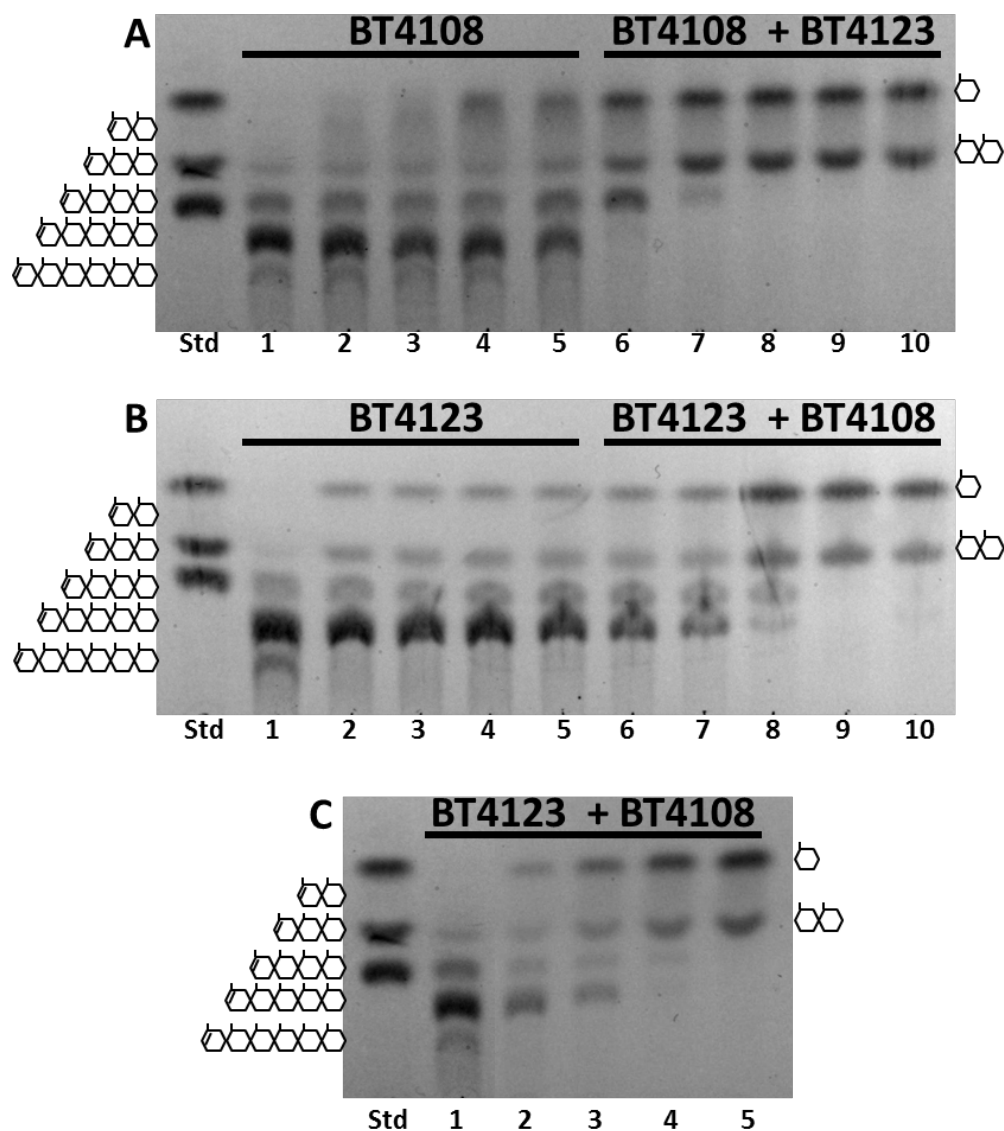


Figure 3.6: TLC analysis of GH105 (BT4108) and GH23 (BT4123) complementary time course digestions of unsaturated oligogalacturonides. (A) BT4108 digestion of uGalA_n time course. Std- GalA, GalA₂, GalA₃. Time points were taken at 0, 5, 60, 180 min and overnight post BT4108 addition, lanes 1-5, respectively. Time points were taken at 0, 5, 10, 30, 60 min and overnight post BT4123 addition to original digest, lanes 6-10, respectively. **(B)** BT4123 digestion of uGalA_n time course. Std- GalA, GalA₂, GalA₃. Time points were taken at 0, 5, 60, 180 min and overnight post BT4123 addition, lanes 1-5, respectively. Time points were taken at 0, 5, 10, 30, 60 min and overnight post BT4108 addition to original digest, lanes 6-10, respectively. **(C)** Co-digestion of uGalA_n with both BT4108 and BT4123. Time points were taken at 0, 5, 60, 180 min and overnight, lanes 1-5, respectively. All digests were prepared with 1μM enzyme, 20 mM sodium acetate pH 5.0 and 1μM spiked enzyme. Carbohydrates to the left (unsaturated) and right (saturated) margins indicate sizes of subsequent bands.

course, displaying the need for an additional activity to further digest the remaining GalA_n.

3.3.2.2 pH optimum— BT4123 is predicted to be a polygalacturonase, and its activity was confirmed using saturated oligogalacturonides (GalA_n) (Figure 3.7). Subsequent analysis at varying pH, to determine pH optimum, was attempted using GalA_n. Aliquots were taken at individual time points and visualized by TLC to determine pH optimum. TLC results of BT4123 GalA_n digests displayed optimal activity between pH 5 and 8, based on the abundance of products (Figure 3.7A-E). This is consistent with a previously reported literature values of GH28 optimum¹²¹.

3.3.2.3 GH28 kinetic analysis— To quantify BT4123 activity, initial velocities for three defined GalA₃ substrate concentrations were obtained by kinetic analysis using a modified Miller's assay¹²⁰. The observed initial velocities were 0.0377, 0.0509 and 0.134 $\mu\text{M s}^{-1}$ for 0.4, 0.5 and 1 mM GalA₃, respectively. Extrapolating a K_m from these initial velocities provides a value of 0.5 mM. This value represent the minimum K_m possible for BT4123, as the addition of further data points with an increased concentration of substrate would increase the V_{max}, directly increasing the K_m. This extrapolated value is consistent with a previously published K_m of 0.34 mM for an exolytic GH28 from the thermophilic bacterium *Thermotoga maritima* acting on GalA₃¹²¹. A full kinetic analysis of the GH28 enzyme was not completed due to substrate cost and difficulties in substrate preparation. Specifically, the charged uronic acid moiety and coordination properties of the substrate make isolating homogeneous substrates difficult. Also, generating reproducible data with the reducing sugar assay is difficult for uronic acid substrates.

(Within the literature there exists a lack of reported success with this technique for uronic acid containing sugars).

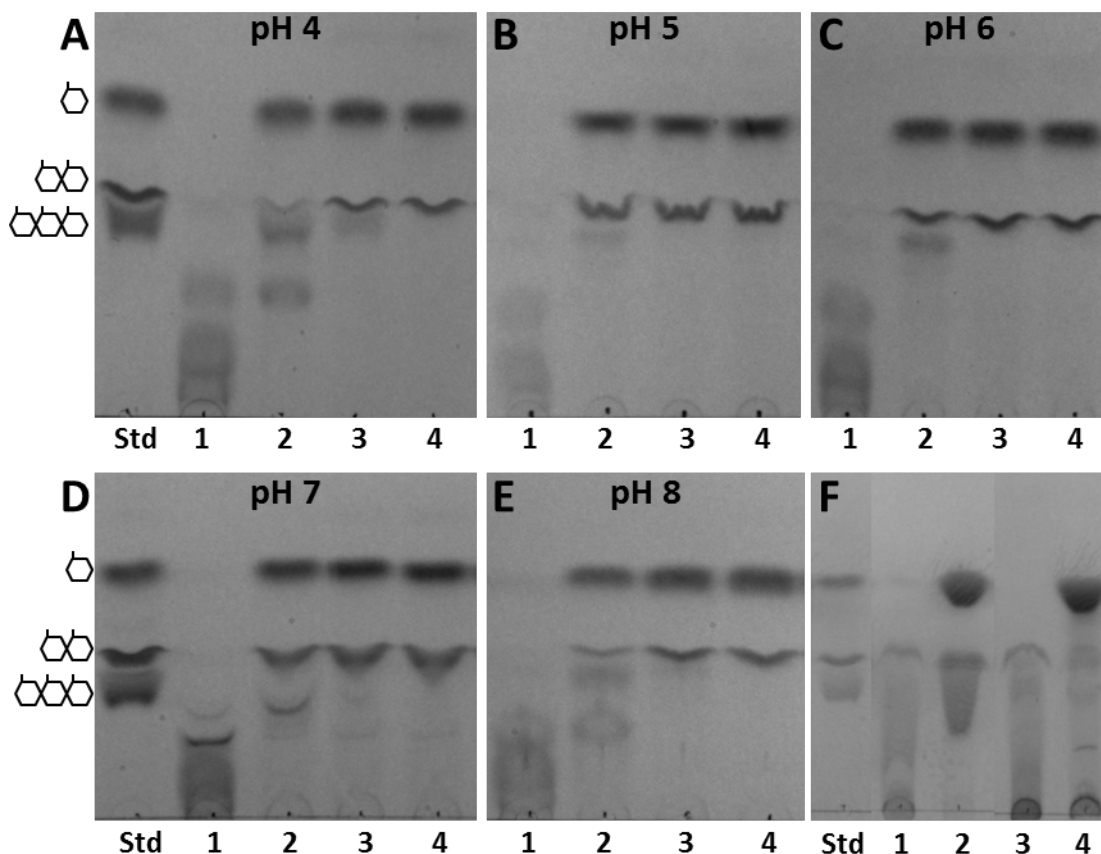


Figure 3.7: TLC analysis of BT4123 (GH28) pH optimum and reducing end assay. TLC analysis of BT4123 ($1 \mu\text{M}$) digestions of GalA_n substrate (1 mg ml^{-1}). Std- GalA, GalA₂ and GalA₃ standards, sizes indicated on left. Lanes 1-4; 0, 5, 10 and 30 minute time points of BT4123 digestion. **(A)** pH 4 **(B)** pH 5 **(C)** pH 6 **(D)** pH7 **(E)** pH8 **(F)** TLC of NaBH_4 reduced and non-reduced GalA₃. Non-reduced GalA₃ (lanes 1-2) or reduced GalA₃ (3-4), were incubated with BT4123 ($1 \mu\text{M}$) (lanes 2 and 4) overnight. Std- GalA, GalA₂ and GalA₃ standards, sizes indicated on left.

3.3.2.4 Product profile— Overnight incubation of HG with BT4123 resulted in some GalA product formation (Figure 3.8B). Based on this activity, we incubated BT4123 overnight with GalA_n and analyzed the products via TLC. At completion, the reaction appears to generate GalA and GalA₂ products (Figure 3.7A-E). To further investigate substrate specificity, digestions using GalA₂ and GalA₃ were attempted with BT4123. GalA₃ was digested to GalA₂ and GalA, and GalA₂ displayed a slow side activity with GalA being produced (Figure 3.5C). HPAEC-PAD analysis confirmed the preferential activity on GalA₃, as compared to the GalA₂ substrate (Figure 3.5F). To determine the role of BT4123 on unsaturated products, BT4123 was tested for activity against uGalA_n. Interestingly, complete digestion of uGalA_n products was not observed (Figure 3.6B). The generation of GalA-sized products was observed; however, larger sized products (> GalA₃) remained. These results were in contrast to the GalA and GalA₂ products observed from BT4123 digestion of GalA_n, as described previously (Figure 3.7A-E). Importantly, these results indicate that BT4123 displays specificity for GalA_n (where $n \geq 3$); however, a side activity on uGalA_n does exist.

3.3.2.5 Reduced GalA₃ digestions— To investigate whether the reducing end of GalA₃ plays a role in substrate recognition by BT4123, the reducing end of GalA₃ was reduced using NaBH₄. Reduced and non-reduced GalA₃ were subject to overnight incubation with BT4123. TLC analysis of the overnight digestions revealed that both reduced and non-reduced GalA₃ substrates were cleaved into the corresponding GalA and GalA₂ products (Figure 3.7F), indicating that the reducing end is not required for recognition or enzymatic activity.

3.3.3 Complementary activities between GH105 and GH28

3.3.3.1 *uGalA_n* BT4123 and BT4108 co-digestions— Based on our individual product profiles, BT4108 removed the unsaturated moiety from the non-reducing end of *uGalA_n* (Figure 3.7A). BT4123 removed GalA from *GalA_n* in an exolytic fashion, and demonstrated very little activity on *GalA₂* and *uGalA_n* (Figure 3.6B). Based on their displayed activities, and as members of PUL75, we sought to investigate a possible relationship between BT4108 and BT4123. Namely, do these enzymes exhibit complementary activities? Three separate digestions were prepared with exclusive production of GalA and *GalA₂* observed upon complete digestion (Figure 3.6). Importantly, complete digestion of the substrates was not observed in the individual time course reactions until after the addition of the second enzyme, highlighting the complementary activities required to depolymerize *uGalA_n* (Figure 3.6).

3.3.3.2 *Matrix digestions with GH105, PL1 and GH28—* Having demonstrated the activities of both GHs within PUL75, we decided to explore the effect of PL1 digestion on HG depolymerisation. We sought to partially reconstitute the HG utilization pathway *in vitro*. We designed a matrix with members of each of the three enzyme families, PL1 (BT4115), GH105 (BT4108) and GH28 (BT4123). See Table 3.1 for reaction contents. In the presence of all three enzymes, HG was deconstructed into GalA and *GalA₂* (Figure 3.8G). Minimal traces of *uGalA₂* can be seen on the TLC, as the digestions were carried out at pH 8, which is much more alkaline than the GH105 pH optimum. Furthermore, these matrix digestions demonstrated the unique role of each enzyme in the deconstruction process, by subsequent digestions lacking one or two of the three enzymes

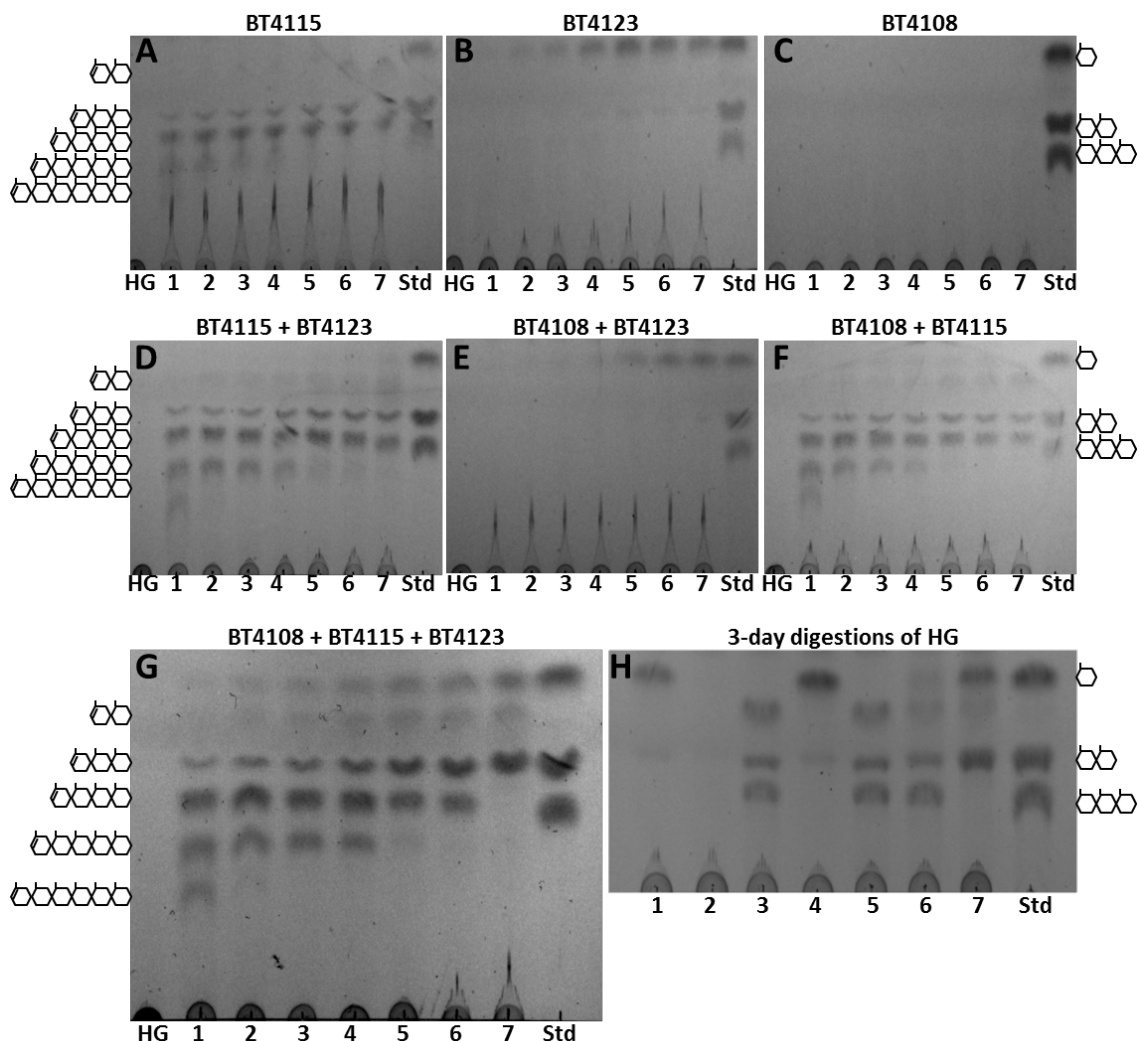


Figure 3.8: Matrix digestions of HG with BT4108, BT4115 and BT4123 (GH105, PL1 and GH28, respectively). All reactions were set up with $1\mu\text{M}$ enzyme as indicated, 20 mM tris pH 8.5 and 0.1 mM CaCl_2 . **(A-C)** Individual HG digestion of BT4115, BT4123 and BT4108, respectively. **(D-F)** HG digestions with co-incubation of BT4115 + BT4108, BT4115 + BT4123 and BT4108 + BT4123, respectively. **(G)** HG digestion with BT4108, BT4115 and BT4123 present. All lanes were loaded with $3\mu\text{l}$ of sample twice, with time points at 0, 5, 15, 30, 60, 180, 300 min and overnight, lanes HG, 1-7, respectively. Std- GalA, GalA₂, GalA₃ standard. Time points were heat treated at 100°C to terminate the reaction, centrifuged to clarify and spotted onto the plate. **(H)** 3-Day digestions of HG with BT4123, BT4108, BT4115, BT4108 + BT4123, BT4115 + BT4123, BT4108 + BT4115 and BT4108 + BT4115 + BT4123, lanes 1-7, respectively. Std- as above. 3-Day aliquots were obtained from above samples A-G, and analyzed in the same manner. Carbohydrates to the left and right margins indicate sizes of visualized products.

(Figures 3.8A-H). BT4115 (PL1) is responsible for the generation of uGalA_n products. In the absence of BT4115, only GalA was produced (Figure 3.9E). BT4123 (GH28) is responsible for the digestion of GalA_n and generates GalA and GalA₂. In the absence of BT4123, GalA_n remains present (Figure 3.8F). Finally, BT4108 (GH105) is required to remove the unsaturated moiety from the uGalA_n, resulting in GalA_n, which is the preferential substrate for BT4123 activity. The matrix digestions revealed a mechanistic requirement for each of the three activities in the complete deconstruction of HG.

3.3.3.3 pH effect on matrix digestions— Due to the range in pH optima between the three enzymes, we sought to investigate the role of pH on product production in our partially reconstituted pathway. Three separate incubations were prepared at varying pHs. Two of the reactions (Figure 3.9) were left to digest at their initial pHs, 8 and 5, respectively. These pHs are the optima for the PL1s and GHs, respectively. The amount of product produced in the pH 8 digestion is far greater than that of the pH 5 (Figure 3.9A vs. Figure 3.9B). This is due to the fact that the vast majority of HG digestion is PL1 dependent, and as such, pH 8 is more favorable to PL1 product generation. Consistently, at pH 5, the PL1 activity (i.e. production of uGalA_n), is at suboptimal pH as determined by the decrease in products generated. In an attempt to tailor the pH for both classes of enzymes, a third incubation with two separate pHs was attempted. Initially pH favored the PL1 activity (pH 8). After overnight digestion, HCl was subsequently added to drop the pH to the GH optimum (pH 5). This change in pH did not seem to have an effect on product generation (Figure 3.9C). Interestingly, distinct pH optima observed for PL1s and GHs may highlight the role of proximal residues in affecting microenvironments, and thus the pKa of catalytic residues¹²².

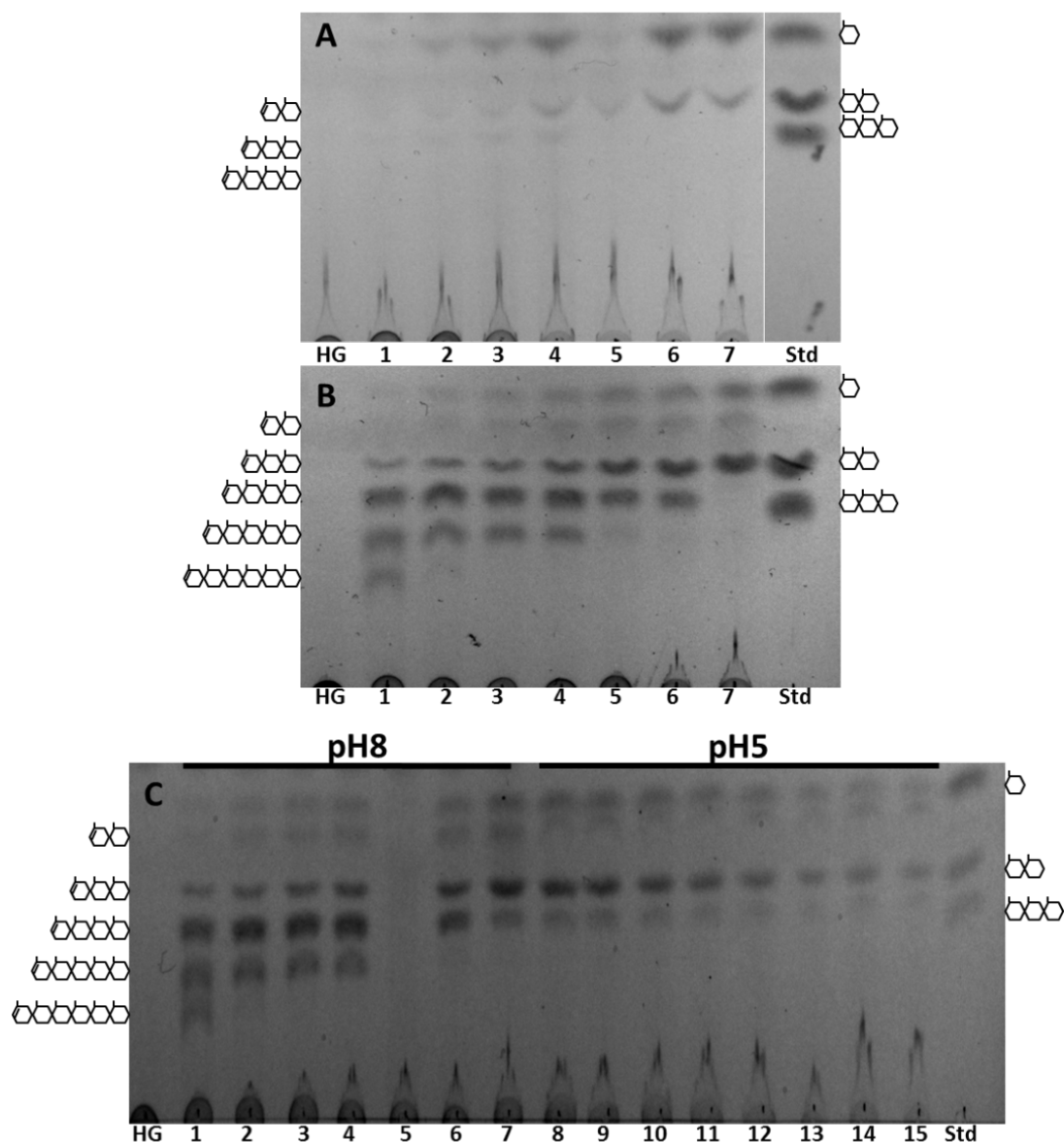


Figure 3.9: pH effect on BT4108, BT4115 and BT4123 matrix digestion of HG (A and B). Time course digestion of HG with BT4108, BT4115 and BT4123. Time points were taken at 0, 5, 15, 30, 60, 180, 300 min and overnight, lanes HG, 1-7, respectively. Std- GalA, GalA₂, GalA₃. **(C).** Time course digestion of HG with BT4108, BT4115 and BT4123. Time points were taken at 0, 5, 15, 30, 60, 180, 300 min and overnight, lanes HG, 1-7, respectively. Std- GalA, GalA₂ and GalA₃. Subsequent time points were taken at 5, 10, 15, 30, 60, 180, 300 min and overnight, lanes 8-15, respectively. Carbohydrates in the margins indicate product sizes. N.B. Panel C-Lane 5 was the result of leaving the reaction at 100°C for an extended period of time. All reactions were set up with 1μM enzyme as indicated, 0.1 mM CaCl₂ and 20 mM sodium acetate pH 5 or tris pH 8, A and B, respectively.

3.4 Discussion

As discussed in chapter 2, the successful characterization of the three PL1s present in PUL75 revealed the initial stages of HG deconstruction by PUL75, namely, the generation of unsaturated oligogalacturonide (uGalA_n) products. However, a large void remained in our understanding of how these uGalA_n products were further deconstructed into the metabolites required for energy production by the bacterium. The successful cloning and recombinant protein production of the gene products BT4108 and BT4123 from PUL75 has enabled the biochemical characterization of the two different glycoside hydrolases (GH105 and GH23, respectively) on uGalA_n PL1 products enabling the partial reconstitution of the PUL75 pathway.

A GH105 homolog of BT4108, YteR from *Bacillus subtilis*, has been previously characterized and displayed activity on unsaturated rhamnogalacturonides (u(Rha-GalA)_n)¹⁰⁶. In this light, we initially screened BT4108 for activity against u(Rha-GalA)_n (data not shown); however, no activity was observed. Subsequent digestions of uGalA_n with BT4108 revealed the removal the unsaturated moiety from the non-reducing end, which is a novel exolytic activity for the GH105 family (Figure 3.5A-B). Further characterization of the pH optimum and kinetic analysis of this novel activity were obtained. BT4108 displayed an acidic pH optimum in the range of 3.6-4.4, with maximal activity between 4.2-4.4 (Figure 3.4A). A previously characterized GH105, YteR¹⁰⁶, identified aspartate (D143) and histidine (H189) as the two key catalytic residues, which are conserved in BT4108 (not shown). Full kinetic analysis of BT4108 (Figure 3.4B) revealed k_{cat} and K_m values of $1.69 \pm 0.07 \text{ s}^{-1}$ and $82 \pm 11 \text{ }\mu\text{M}$ (Table 3.2). YteR,

revealed a k_{cat} and K_m of $0.280 \pm 0.011 \text{ s}^{-1}$ and $100 \pm 14 \text{ }\mu\text{M}$, respectively¹⁰⁶. BT4108 has roughly a 6-fold increase in catalytic turnover and a 7-fold increase in catalytic efficiency (k_{cat}/K_m) compared to its family member YteR.

TLC analysis of BT4108 overnight digestions confirmed exolytic activity on a range of uGalA_n substrates (Figure 3.5A-B). Resolution and successful identification of the generated products with TLC analysis was challenging, as GalA_n and uGalA_{n+1} ($n \geq 2$) have nearly identical mobility during TLC analysis. In this regard, HPAEC-PAD was required to correctly identify the generated products. uGalA_n products are retained longer than GalA_n in HPAEC-PAD analysis due to the structural and hydrophobic changes of the molecule producing greater resolution. The HPAEC-PAD analysis (Figure 3.6D-E) confirmed the removal of the unsaturated moiety from the non-reducing end, resulting in the generation of a linear 5-keto-4-deoxyuronate (DKI) product which is difficult to visualize on TLC and to resolve in HPAEC-PAD and remaining saturated saccharide(s) (Figures 3.5D-E). Based upon these product profiles the BT4108 active site contains a minimum of 2 subsites (-1 and +1) for activity (Figure 3.10A), as the enzyme is active on uGalA₂. Based on the subsite nomenclature (Figure 1.4), the -1 subsite is populated by the unsaturated moiety (uGalA) and a GalA is positioned in the +1 subsite. Further interactions that may exist between enzyme subsites and the substrate in positive direction ($>+1$) and any implications of these interactions for enzyme activity remain to be determined.

Time course digestions revealed that after exolytic activity, removing the unsaturated moiety, the remaining saturated oligosaccharide (GalA_n) products are not further

digested by BT4108 (Figure 3.6A). This fact highlights the need of a complementary enzyme to further cleave these products into substrates that can act as energy generating molecules for the bacterium.

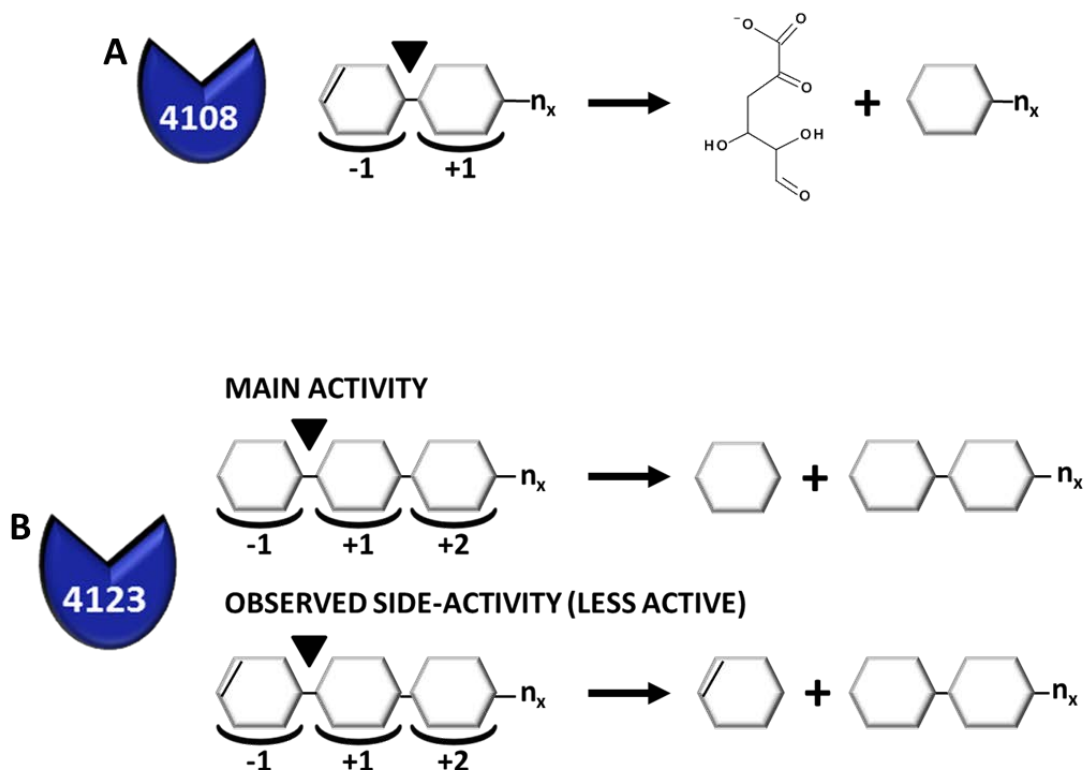


Figure 3.10: Proposed subsite orientations for (A) exolytic GH105 (BT4108) and (B) exolytic GH28 (BT4123)

As the only predicted polygalacturonase in PUL75, biochemical characterization of BT4123 was attempted to elucidate its role in the deconstruction pathway. BT4123 digestion products are saturated, and therefore, cannot be directly quantified via spectrophotometric detection. In this regard, we employed a TLC approach to determine the pH optimum of BT4123. Time courses, with pH ranging from 4-8, were performed (Figure 3.7A-E). Based on the analysis and comparison of the products present at given time points, we determined BT4123 was active over the entire pH range, with an optimum of pH 5. Previously characterized GH28s, revealed a catalytic triad consisting of aspartate residues, which is consistent with the determined pH optimum for BT4123¹²¹. A variation of the reducing end assay¹²⁰ was used to determine some kinetic parameters of BT4123. At this time, a full kinetic analysis has yet to be completed; however, at select substrate concentrations (0.4, 0.5 and 1 mM) initial rates of 0.0377, 0.0509 and 0.134 $\mu\text{M s}^{-1}$ were obtained. Extrapolation based on these results suggests that the rate of hydrolysis by BT4123 may be within 10-15% of a previously reported GH28 family member¹²¹. In an attempt to gain insight into substrate orientation within the active site and if the role of the reducing end was a recognition determinant, we incubated BT4123 with a reduced and non-reduced GalA₃. The overnight digestion resulted in the generation of GalA and GalA₂ products, consistent with our previous analysis (Figure 3.7F). This suggests that the reducing end does not play a significant role in enzyme-substrate recognition. Interestingly, BT4123 displayed a side activity on uGalA_n substrates. A time course of BT4123 digestion of uGalA_n revealed the generation of a GalA product, however the remaining larger oligosaccharides were not further hydrolyzed (Figure 3.6B). This activity demonstrated that the non-reducing end has an

increased role in enzyme-substrate recognition. In this regard, we propose that the BT4123 active site contains a minimum of 3 subsites (-1, +1 and +2) (Figure 3.10B) based on its preference of GalA_n (where $n \geq 3$). Furthermore, the -1 subsite is preferentially populated with a saturated moiety for complete digestion. However, activity can still be observed, albeit at a reduced rate, when an unsaturated moiety is positioned in this site.

Interestingly, both of the GH product profiles demonstrated that individually, the enzymes are unable to completely hydrolyze uGalA_n, products from upstream PL1 activity, into GalA. However, there appears to be a complementary nature between these two GHs. In order to validate this integrated relationship, we carried out time course digestions as previously described with mixtures of BT4108, BT4123 and uGalA_n. As expected, individually, the enzymes were unable to completely hydrolyze the products. When the complementary enzyme was spiked into the overnight digestion, the generation of monosaccharides and disaccharides (GalA and GalA₂) was observed (Figure 3.6A-B). Similarly, a co-digestion of uGalA_n with both enzymes simultaneously revealed the generation of GalA and GalA₂ products (Figure 3.6C). The discovery and demonstration of these complementary enzymes is key to our understanding of the systematic deconstruction of HG by PUL75 gene products, namely, in how the large pool of uGalA_n substrates were further hydrolyzed.

As we now had determined the individual roles of each the enzymes in the deconstruction pathway, we sought to attempt the partial reconstitution of the HG utilization pathway *in vitro* by creating a digestion matrix involving three PUL75 enzymes; BT4108, BT4115

and BT4123. BT4115 was preferentially chosen over the other two PL1s based on its product profile (Figure 2.6). In this light, we analyzed time courses of the matrix via TLC to display product profiles and gain insights into HG deconstruction by this abbreviated pathway (Figure 3.8A-H). Individual HG digestions confirmed our previously reported activities and product profiles (Figure 3.8A-C), with BT4115 producing uGalA₂, uGalA₃ and uGalA₄ products; BT4123 producing GalA; and BT4108 displaying no activity. Co-digestions with two of the three enzymes revealed the complementary mechanism of all three enzyme activities. When no BT4108 is present (Figure 3.8D), uGalA_n products are readily produced; however, BT4123 is not preferentially active on unsaturated substrates, and we do not see complete digestion. When BT4123 is absent (Figure 3.8F) the unsaturated moiety from uGalA_n is hydrolyzed but saturated oligosaccharides (GalA_n) remain. When BT4115 is absent, GalA accumulates slowly due to BT4123 activity; however the amount of polymerized HG remaining is significant, due to the lack of endolytic depolymerase (Figure 3.8E). The presence of all three enzymes deconstructed HG into monosaccharide (GalA) and disaccharide (GalA₂), respectively (Figure 3.8G).

We performed further digestions with all three enzymes present at two pHs, one near the optimum of the PLs (pH 8), and one closer to those of the GHs (pH 5) (Figure 3.9A & B). Both digestions displayed the characteristic GalA and GalA₂ final products; however, at pH 5 the PL1s displayed suboptimal activity, and the amount of products produced was considerably lesser. These results highlight the importance of an alkaline pH for PL1 activity. Though the observed pH optima of BT4108 and BT4123 were acidic (pH 4.2 and pH 5), the proximal amino acid residues (localized microenvironment) may play a

role in ‘fine-tuning’ the pKa values of the catalytic residues, providing a broader pH optima range for the pathway that is consistent with the physiological conditions present in the distal gut.

To completely deconstruct HG into GalA and GalA₂, we require at least three defined activities, endolytic PLs (BT4115), and two different exolytic GHs (BT4108 and BT4123). We designed a time course assay that began at pH 8, to favor the initial endolytic PL activity, with a subsequent adjustment to pH 5, to favor the complementary GH activities (Figure 3.9C). Consistent with previous results, we see a large quantity of uGalA_n production, with some, albeit quite small, GH activity being observed. Upon altering the pH, we see a dramatic increase in the GH activity, consistent with their pH optima, and the generation of the GalA and GalA₂ end-products. This experiment highlights the pH range required for optimal activity of the three enzyme families. A subsequent experiment using a more neutral pH (6.5-7) may be of interest as the addition of concentrated HCl may have had a negative effect on localized protein, thus affecting the overall digestion. Despite this observation we have provided compelling evidence that the PUL75 pathway can be reconstituted *in vitro*, and in the process have described a new activity for a GH105 family member.

3.5 Conclusion

The successful production and purification of recombinant BT4108 and BT4123 enabled the biochemical characterization of these two enzymes. BT4108 (GH105) revealed a new activity on uGalA_n substrates as it removes the unsaturated moiety from the non-reducing end of uGalA_n (where $n \geq 2$). Kinetic analysis of this enzyme revealed a k_{cat} and K_m of $1.69 \pm 0.07 \text{ s}^{-1}$ and $82 \pm 11 \text{ }\mu\text{M}$, respectively and pH optimum in the range 4-4.2. BT4123 (GH28) was confirmed as an exolytic polygalacturonase, preferentially active on the non-reducing end of substrates with a degree of polymerization (DP) of $n \geq 3$. Partial kinetic analysis revealed an optimal K_m near 0.5 mM, however complete kinetic analysis remains unavailable at this time. The GH28 and GH105 displayed complementary activities towards the uGalA_n lyase products, resulting in the production of GalA and GalA₂, the final products in HG metabolism by PUL75. The digestion of HG with BT4108, BT4115 and BT4123 (GH105, PL1 and GH23, respectively), demonstrated the successful partial reconstitution of the PUL75 HG utilization pathway *in vitro*.

Chapter 4-

Characterization of the PUL75 Sus-like proteins: insights into HG binding and affinity

4.1 Introduction

In previous chapters, we described the activities of how PLs and GHs, two different classes of enzymes, produce uGalA_n and GalA_n from HG. Interestingly, a large number of these activities occur within the periplasm of *Bacteroides* spp, therefore transport of the HG or its fragments into the bacterium is essential pathway utilization (Table 4.1). In other PULs the recruitment and effective transport into the bacterium is performed by SusD-like and susE-like outer membrane anchored proteins. These proteins are named for their homology to the well characterized and defined starch utalization system (Sus) of *B. thetaiotaomicron*.

4.1.1 Structural highlights— Crystal structures of the SusD (PDB ID: 3CKC)⁸¹ and SusE (PDB ID: 4FE9)⁸³ proteins have been solved and revealed two different architectures: SusD displays an α -helical fold (Figure 4.1) and SusE a β -sandwich fold (Figure 4.2). Interestingly, the SusD structure also revealed the presence of a number of tetratricopeptide repeats (TPRs), associated with protein-protein interactions¹²³, that were not predicted in the sequence analysis of SusD. The presence of TPRs in SusD suggests that these proteins may form part of a larger protein complex, which has implications for higher order structures on the outer membrane and mechanisms of carbohydrate recognition and binding affinity.

Table 4.1: PUL75 gene product IDs and localization, based on SignalP 4.0 and InterProScan 4.8 bioinformatic analyses.

Gene	Family	Signal Peptide	Predicted Localization	Other predicted features
<i>BT4108</i>	GH105	(1-24) LSA-QQ	Periplasmic	Non-cytoplasmic
<i>BT4112</i>	SusE-like	(1-23) FTA-CE	Extracellular	Fibronectin Domains (FN ₃)
<i>BT4113</i>	SusD-like	(1-22) LMT-SC	Unknown	TPRs
<i>BT4115</i>	PL1	(1-20) IQT-NT	Unknown	N/A
<i>BT4116</i>	PL1	(1-21) ACS-EE	Unknown	Lipoprotein
<i>BT4119</i>	PL1	None Predicted	Unknown	Transmembrane Region
<i>BT4120</i>	SusE-like	(1-22) LMT-SC	Unknown	Non-cytoplasmic
<i>BT4122</i>	SusD-like	None Predicted	Outermembrane	TPRs
<i>BT4123</i>	GH28	None Predicted	Extracellular	N/A

(1-2X) indicates # of amino acids in predicted signal peptide sequence.

XXX-XX indicates predicted cleavage site of predicted signal peptide.

N/A- not available

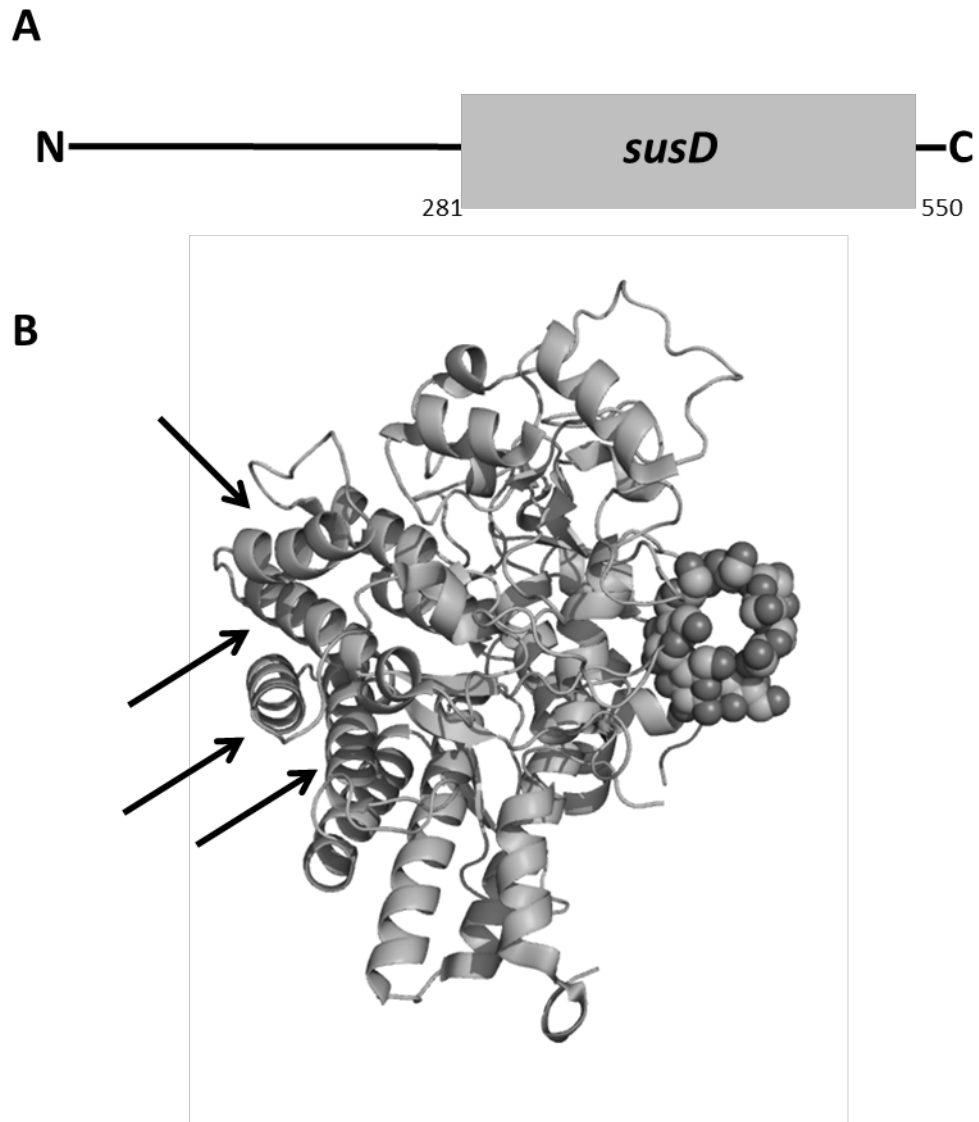


Figure 4.1: SusD in complex with maltoheptaose. (A) Modular structure of *susD*. N and C denote N- and C- termini and numerical values indicate domain boundaries (B) Cartoon representation of SusD crystal structure (PDB ID: 3CK9). Crystal structure revealed the α -helical fold and the endo-type recognition between the protein and 3-D structure of the ligand. α -helices found to contain tetratricopeptide repeats (TPRs) are indicated by the black arrows.

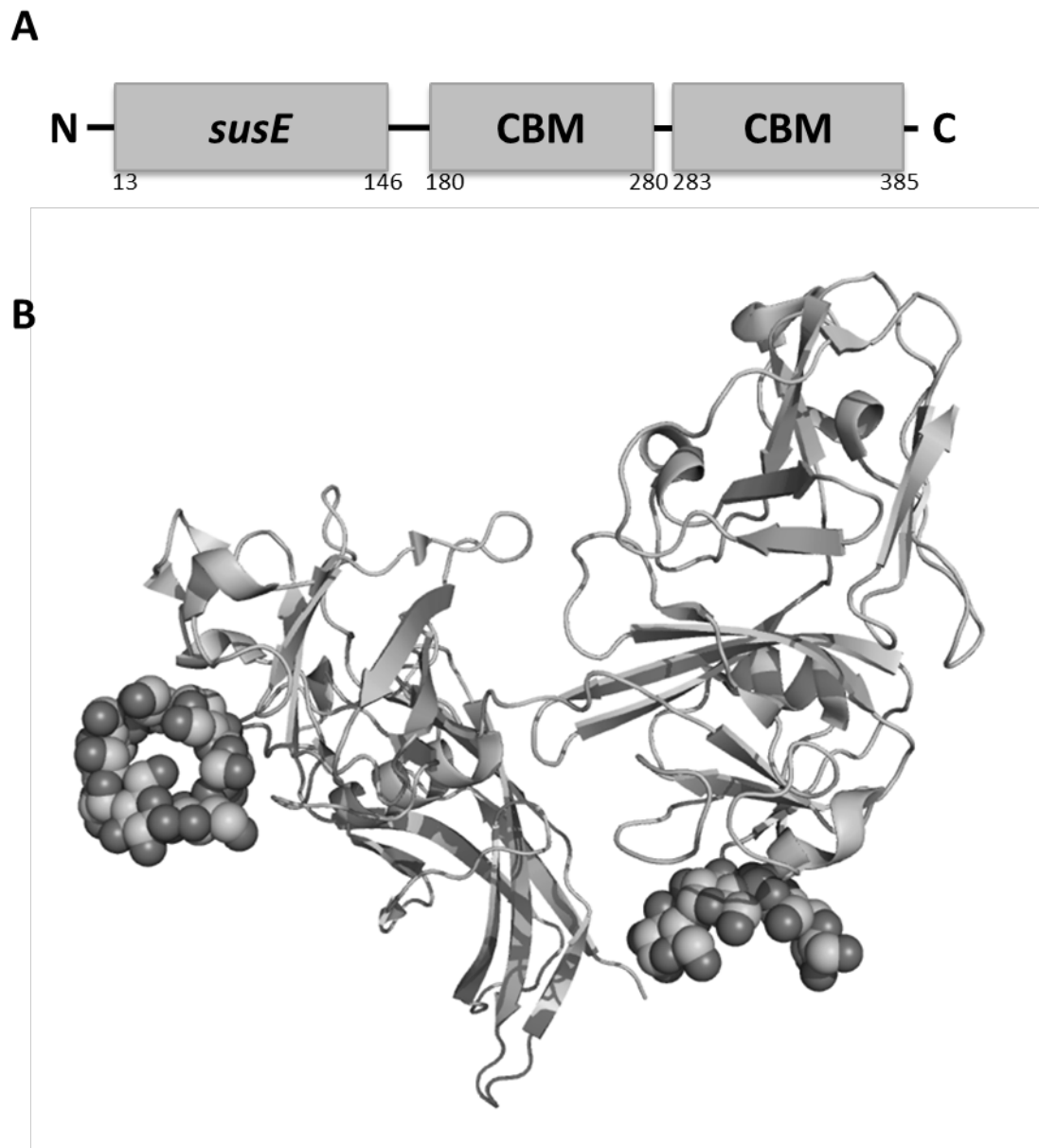


Figure 4.2: SusE in complex with maltoheptaose and maltotetraose. (A) Modular structure of *susE*. N and C denote N- and C- termini and numerical values indicate domain boundaries. CBM- Carbohydrate Binding Module. **(B)** Cartoon representation of SusE crystal structure (PDB ID: 4FE9). Crystal structure revealed the β -sandwich fold and the presence of two carbohydrate binding modules (CBMs) and the endo-type recognition between the protein and 3-D structure of the ligands.

Substrate recognition in SusD is mediated by hydrophobic interactions with three key residues, W98, W320 and Y296. Interestingly, these aromatics create a concave surface, referred to as the aromatic cradle¹²⁴, that complements the 3-D structure of the helical substrate. As such, ligand recognition in the SusD protein is said to be more endo-specific in nature, as opposed to other exo-specific recognition types⁷⁸. A number of SusD-like proteins structures have been solved with each structure possessing the α -helical fold; however, in most cases the target ligand has yet to be identified, thus insights into the molecular determinants of binding are lacking.

In contrast to the single binding domain of the SusD protein, structural analysis of the SusE protein revealed two binding domains linked by a tethering region which was not modeled in the crystal structure⁸³. To date, most starch associated binding domains have been found in proteins that are amylases. As such, the crystal structure of susE, which contains no amylase, was the first non-enzymatic starch binding domain reported⁸³. These tandem binding domains appear to provide an avidity effect, as displayed by ITC analysis with various mutant SusE constructs⁸³. This observed avidity effect is the result of an initial protein-ligand interaction, increasing the proximity of bound ligand to the remaining binding sites with a negligible entropic penalty, thereby increasing the likelihood of a subsequent interaction. There are a number of residues involved in substrate recognition, with W296 and W336 involved in hydrophobic interactions. These hydrophobic interactions are further classified as CH/ π interactions whereby the apolar face of the carbohydrate interacts with the π orbitals of aromatic amino acid residues¹²⁵. In addition to the CH/ π interactions described above, hydrogen bonding interactions between R326 and R350, and hydroxyl groups on the ligand were observed. Similar to

SusD, SusE substrate recognition is also endo-specific in nature, as it recognizes the contoured 3-D structure of the ligand chain.

4.1.2 Sus-like homologs— Each of the 88 PULs identified from *B. thetaiotaomicron* contain a homolog of the SusC and SusD proteins first characterized in PUL66; the starch utilization system (sus). Interestingly, PUL75 contains two copies of the SusC-like and SusD-like proteins (BT4114 and BT4121 and BT4113 and BT4122, respectively). Though these two homologs are present in all of the 88 identified PULs, the presence of two copies of each is not common among PULs and remains poorly understood. In addition to these two copies of the SusC/SusD like proteins, PUL75 contained two large open reading frames (ORFs, BT4112 and BT4120) proximally located to the previously identified Sus-like genes. These ORFs could represent other members of the Sus-like protein family (susE), that have yet to be identified or characterized. In this light, we sought to clone and to produce recombinantly the two copies of each of the SusD gene products (*BT4113* and *BT4122*) and ORF gene products (*BT4112* and *BT4120*) from PUL75 for characterization in an attempt to elucidate their roles within HG recruitment, transport, and utilization.

4.2 Materials and Methods

4.2.1 Recombinant protein production— The coding sequences of the two PUL75 ORFs *BT4112* and *BT4120* were amplified from *B. thetaiotaomicron* genomic DNA using engineered primers (Appendix 1). Amplicons were restricted with *NheI* (5') and *XhoI* (3') enzymes and ligated into pET28a (Novagen, Cat #69864-3) or *NcoI* (5') and *XhoI* (3') pET32a (Novagen, Cat #69015-3) expression vectors with complementary ends to create the expression plasmids pET28-BT4112 and pET28-BT4120. These constructs contain N-terminal 6x histidine cleavable tags for purification by IMAC. All expression plasmids were individually transformed into BL21 (DE3) (EMD Millipore, Cat #69450-3, Karmstadt, Germany) cells for protein overexpression. These transformants were used to inoculate 1 l of LB broth containing kanamycin (50 $\mu\text{g ml}^{-1}$: pET28a) or ampicillin (50 $\mu\text{g ml}^{-1}$: pET32a). Cultures were grown at 37°C to an OD₆₀₀ of 0.8-1.0, cooled to 16°C and induced with 0.2 mM IPTG (Cells were continuously shaken at 180 rpm during the growth and induction process). The following day the cells were harvested by centrifugation at 6,500 x *g* for 10 min at 4°C. The cell pellet was re-suspended in sucrose solution [50 mM tris-HCl (pH 8.0), 25% sucrose with 10 mg l⁻¹ lysozyme] and chemically lysed at ambient temperature with 2 volumes of lysis buffer [1% deoxycholate, 1% triton X-100, 20 mM tris-HCl (pH 7.0), and 100 mM NaCl] with 200 $\mu\text{g l}^{-1}$ DNase. The lysate was clarified by centrifugation at 17,500 x *g* for 45 min at 4°C and filtered using a 0.45 μM low protein binding syringe filter (Pall Acrodisc 25mm). This soluble protein fraction was applied to a nickel (II) nitrilotriacetic acid (Ni²⁺-NTA) sepharose support (GE Healthcare, Cat#: 17-5318-06) and the immobilized protein was washed with 500 mM NaCl, 20 mM tris-HCl (pH 8.5) to decrease non-

specific binding. A stepwise imidazole gradient (5 mM-500 mM) was used to elute the target protein. Samples were visualized by SDS-PAGE, and fractions with appreciable amounts of target protein were pooled and dialyzed overnight against 4 l of 20 mM tris (pH 8.0), with two buffer exchanges, using a dialysis membrane (Spectrum, Biotech CE) with 5,000 MW cutoff. The following day protein concentration was determined at OD_{280nm} using extinction coefficients of 43,360 M⁻¹ cm⁻¹ (0.732 mg ml⁻¹) and 65,040 M⁻¹ cm⁻¹ (1.065 mg ml⁻¹) for BT4112 and BT4120, respectively, as calculated by the ProtParam analysis tool ¹¹⁹.

BT4113 and BT4122 plasmids were obtained from our collaborator, Dr. Harry Gilbert (University of Newcastle-upon-Tyne). Purified plasmid was transformed into BL21 cells and soluble protein was produced and purified as described above. Protein concentration was determined at OD_{280nm} using extinction coefficients of 71,003 M⁻¹ cm⁻¹ (2.111 mg ml⁻¹) and 158,640 M⁻¹ cm⁻¹ (2.049 mg ml⁻¹) for BT4113 and BT4122, respectively, as calculated by the ProtParam analysis tool ¹¹⁹.

4.2.2 Affinity gel electrophoresis (AGE)— Native acrylamide gels were prepared with 10% stacking and 4% separating. Stacking gels included 2% soluble carbohydrate, that became embedded into the stacking gel matrix. The carbohydrates selected were HG, RGI, PG and starch. A native gel (negative control) was also prepared where distilled water was substituted in place of soluble carbohydrate. Bovine Serum Albumin (BSA) acted as a negative control for protein carbohydrate interaction and a Carbohydrate Binding Module (CBM) 32 protein with a defined affinity for HG¹²⁶ acted as a positive control. Proteins were loaded onto the gels and run at 150V for 3 h at 4°C¹²⁷. Proteins

were visualized with Coomassie Brilliant Blue G-250 and migration of the samples in the different carbohydrate gels was compared to the migration in the native gel (negative control).

4.2.3 Isothermal titration microcalorimetry (ITC)— Protein samples were produced as described above and exhaustively dialyzed for a minimum of 4 h (minimum 3 buffer exchanges) against 20 mM tris (pH 8.0) using a 5,000 MW cut-off membrane (Spectrum, Biotech CE). Protein was concentrated in a stirred Amicon 8200 filtration device (pressurized N₂ gas with 5,000 MW cut-off membrane (Pall)) to the desired concentration as determined by A₂₈₀ using the corresponding extinction coefficients. Filtrate was saved and used to dilute ligand. Ligands were obtained by large-scale digests and column chromatography (as discussed in section 2.2.8) and diluted to a final concentration of 1 mM using the saved filtrate.

Instrument and sample preparation of the VP-ITC was carried out as outlined in the VP-ITC operator's manual¹²⁸. This includes at least one 5 minute degassing at experimental temperature of both ligand and acceptor, flushing of acceptor cell with buffer (minimum 3X) and charging syringe with ligand (minimum 2X). Once sample preparation and set-up was complete, the program was initialized and ran until completion. Program parameters were as follows: 25°C, 450 rpm stirring speed, 29 injections (1st injection 2 µl, and 10 µl for remaining 28 injections), 60 second initial delay with 210 second delay between remaining injections. Raw ITC data was analyzed using Origin™ 7.0 to obtain the data presented in Tables 4.2 and 4.3 . All reactions were run in triplicate.

In subsequent ITC trials the polymerized HG (Megazyme) acted as acceptor and the protein (BT4112 and BT4120) was the titrant. The same protocol was followed as above; however, the data treatment differed in that the concentration of carbohydrate was converted to coincide with the number of GalA monosaccharides present. This enables size of the 'binding footprint' on the HG chain to be determined for one protein molecule.

4.3 Results

4.3.1 BT4112

4.3.1.1 Cloning, protein expression and purification— The coding sequence of BT4112 was cloned into the pET28a *E. coli* expression vector in order to produce recombinant protein for characterization. Gene sequencing confirmed no mutations in the construct. BT4112 was expressed as an approximately 59 kDa soluble protein and purified by Immobilized Metal Affinity Chromatography (IMAC). BT4112 fractions were pooled based on SDS-PAGE analysis (Figure 4.3C) and dialyzed. Approximately 16 mg BT4112 was obtained per 1 l culture.

4.3.1.2 Oligosaccharide production, separation and purification— Unsaturated oligogalacturonides [uGalA_n where n is degree of oligomerization] were obtained as described in 2.2.8. In order to obtain saturated oligogalacturonides (GalA_n), small scale digests of HG with an endo-polygalacturonase (*Rhizopus* spp., Sigma Cat# 9032-75-1) were performed in sodium acetate (pH 4.0) to produce GalA_n ($n=2-7$ sugars in length). Incubation times of the small scale digests were adjusted to produce saturated oligosaccharides 2-7 sugars in length. Digestions were subsequently scaled up 500 times to produce large quantities of these saturated oligosaccharides. The subsequent separation was as outlined in section 2.2.8. (Figure 4.6D-C)

4.3.1.3 Affinity gel electrophoresis (AGE)— BT4112 was subsequently screened for binding to select carbohydrates by AGE. The analysis indicated the presence of a significant shift in the mobility of BT4112 in the presence of HG (Figure 4.4). A slight

shift was also observed when in the presence of RG-I and PG, suggesting a weak interaction.

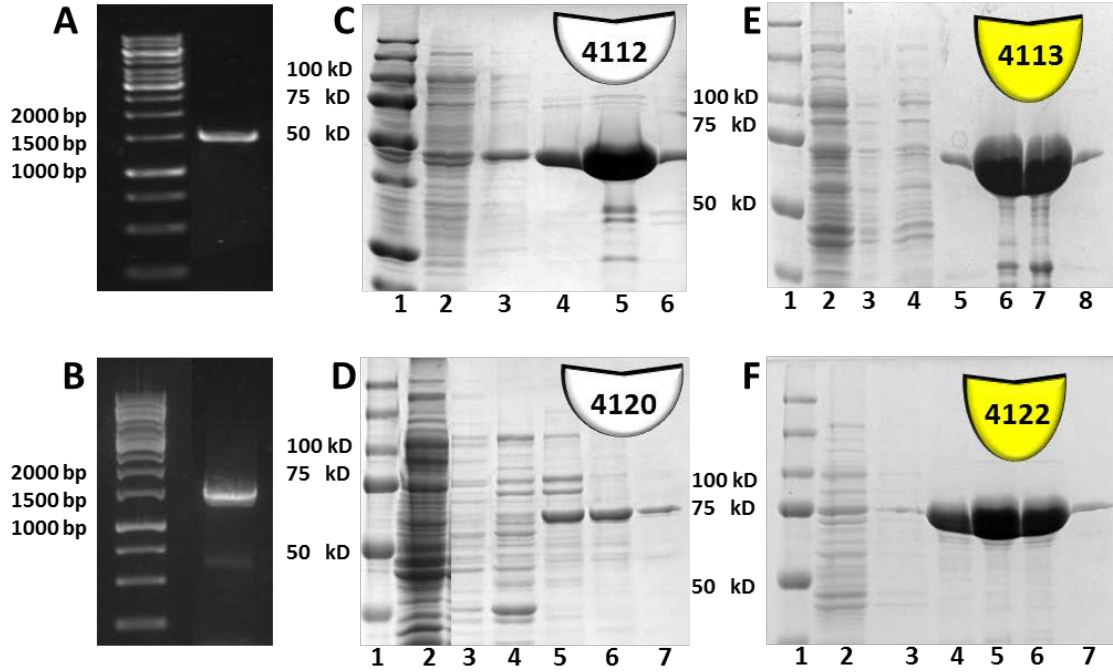


Figure 4.3: Cloning and recombinant protein purification of the two SusD-like and susE-like proteins from PUL75. (A-B) Agarose Gel (1%) used to visualize the PCR product for each of the respective ORF genes, *BT4112* and *BT4120*, respectively. The MWM is on the left hand side of the gel, and the PCR product is on the right. (C) SDS-PAGE analysis of *BT4112* (pET28a) IMAC fractions. Lanes 1-6; MWM, wash, 5 mM, 10 mM, 100 mM and 500 mM elution, respectively. (D) SDS-PAGE analysis of *BT4120* (pET28a) IMAC fractions. Lanes 1-7; MWM, cell lysate, wash, 5 mM, 10 mM, 100 mM and 500 mM elution, respectively. (E) SDS-PAGE analysis of *BT4113* (pET28a) IMAC fractions. Lanes 1-8; MWM, cell lysate, wash, wash, 5 mM, 10 mM, 100 mM and 500 mM elution, respectively. (F) SDS-PAGE analysis of *BT4122* (pET28a) IMAC fractions. Lanes 1-7; MWM, cell lysate, wash, 5 mM, 10 mM, 100 mM and 500 mM elution, respectively.

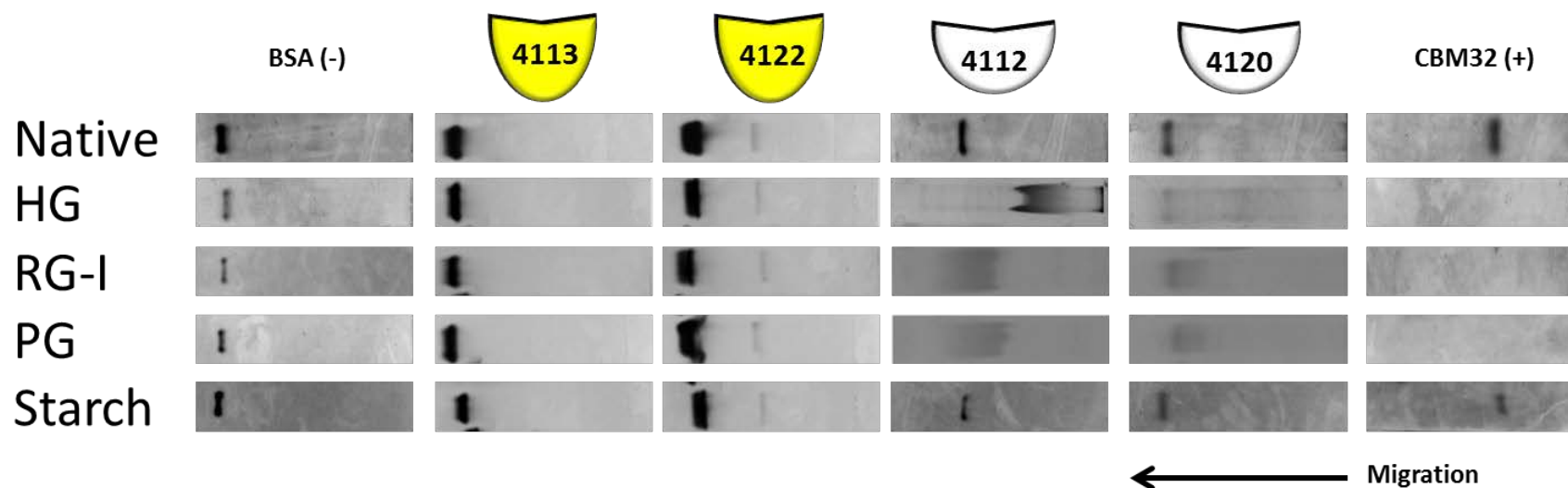


Figure 4.4: AGE carbohydrate analysis. 1% carbohydrate gels [native (negative control), HG, RG-I, PG and starch] were loaded with indicated proteins; BSA (negative control), BT4113, BT4122, BT4112, BT4120 and CBM32 (positive control). The native gel acts as a reference to determine shifts in migration. Migration direction is indicated by the arrow at the bottom of the figure. Equal volumes of sample and non-denaturing buffer were loaded onto each gel lane. A strong interaction is indicated by a distinct shift in migration, whereas as weak interaction is associated with a smear or loss of visualized protein in the lane.

4.3.1.4 ITC binding assays— In an attempt to quantify the affinity between BT4112 and HG, ITC analysis was employed (Figure 4.5A). The binding affinity between BT4112 and HG was observed with a value of $60 \pm 9 \times 10^3$ M. In order to determine the binding ‘footprint’ present on BT4112, the donor and acceptor roles in the ITC were reversed, with protein being titrated into ligand. This provides an n value of 0.058, which equates to an average footprint of approximately 17 monosaccharides for one binding site. Based on the HG hexasaccharide extracted from the PL1 structure of *E. chrysanthemi* EC16 (PDB ID: 2EWE), and a crystallized HG structure¹²⁹; one GalA monosaccharide has a length of 3.77 Å. Using this data, 17 monosaccharides would have a length of ~ 64 Å (Figure 4.5B). Full ITC parameters can be found in Table 4.2.

Table 4.2: ITC binding parameters obtained for BT4112 and BT4120

Protein	Ligand	n	K_a ($10^3 M^{-1}$)	ΔH (kcal mol ⁻¹)	ΔS (kcal mol ⁻¹)	K_d (μM)
BT4112 susE-like	HG	0.058 ± 0.003	60 ± 9	-11.4 ± 0.5	-4.9 ± 0.6	16.8
BT4120 susE-like	N.B.	N.B.	N.B.	N.B.	N.B.	N.B.

N.B. – No binding detected

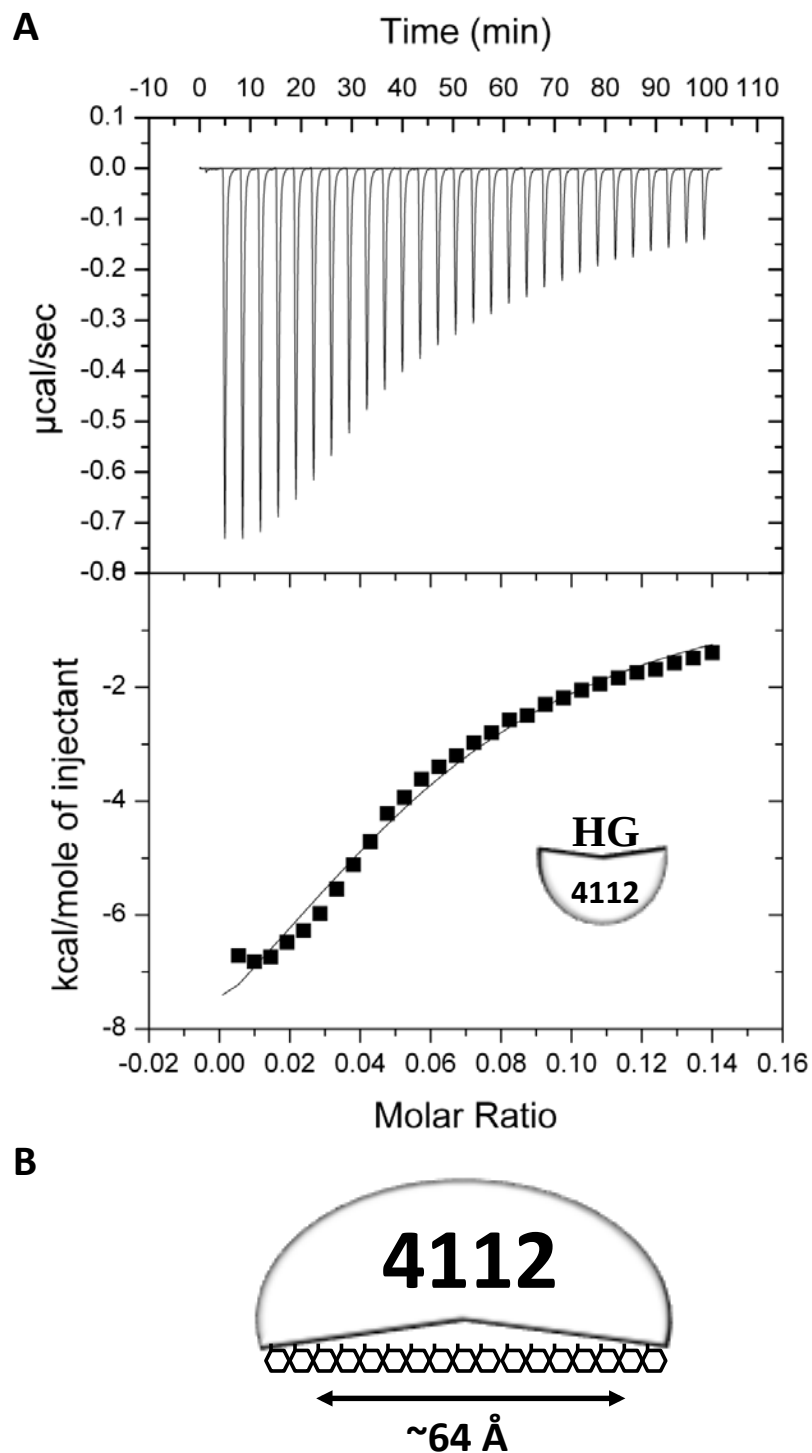


Figure 4.5: ITC binding isotherm of BT4112 (SusE-like) solute binding protein and HG. (A) Top panel- Observed heat trace for each injection of the titration. Lower panel- Enthalpies of binding fit to a one site binding model. **(B)** Cartoon schematic of BT4112 binding foot-print, based on a one-site model.

4.3.2 BT4120

4.3.2.1 Cloning, protein expression and purification— The coding sequence of BT4120, with the predicted signal peptide removed, was cloned into the pET28a *E. coli* expression vector in order to produce recombinant protein for characterization. Gene sequencing confirmed no mutations in the construct. BT4120 was expressed as an approximately 61 kDa soluble protein and purified by IMAC. BT4120 fractions were pooled based on SDS-PAGE analysis (Figure 4.3D) and dialyzed. Approximately 5 mg BT4120 was obtained from 1 l of culture, as determined by A_{280} .

4.3.2.2 Affinity gel electrophoresis (AGE)— BT4120 was screened for binding to HG, RG-I and PG by AGE. This analysis indicated the presence of slight shifts in mobility or smears in the lanes containing HG, RG-I and PG, when compared to the native gel (Figure 4.4). These slight shifts, or smears, are indicative of a very weak interaction.

4.3.2.3 ITC binding assays— In an attempt to quantify BT4120 affinity with HG, ITC was employed in a similar fashion as above (4.3.1.5); however, no affinity was detected (Table 4.2).

4.3.3 BT4113

4.3.3.1 Cloning, protein expression and purification— The coding sequence of BT4113 cloned into the pET28a *E. coli* expression vector was obtained from Dr. Harry Gilbert. BT4113 was expressed as an approximately 71 kDa soluble protein and purified by IMAC. BT4113 fractions were pooled based on SDS-PAGE analysis (Figure 4.3E) and

dialyzed. Approximately 10 mg of BT4113 was obtained from 1 l of culture as determined by A_{280} .

4.3.3.2 Affinity Gel Electrophoresis (AGE)— Based on its homology to SusD, BT4113 was screened for HG, RG-I, and PG binding by AGE. There was no significant shift detected in the presence of any of the screened polysaccharides (Figure 4.4).

4.3.3.3 ITC binding assays— Subsequent ITC analysis was attempted with HG and RG-I polysaccharides, and their constituent monosaccharides, GalA and Rha. BT4113 displayed no affinity to either the mono- or polysaccharides. Based on our understanding of the SusD role in the Sus system, we further attempted ITC analysis with both unsaturated and saturated oligogalacturonides obtained from large scale HG digestions (see sections 2.2.8 and 4.3.1.5). BT4113 displayed affinity for both saturated and unsaturated oligogalacturonides with a minimum length of 4 sugars (Figure 4.6). The affinities for the saturated and unsaturated oligogalacturonides were $2.2 \pm 0.2 \times 10^3$ and $5.3 \pm 1.3 \times 10^3 \text{ M}^{-1}$, respectively. Both values were obtained with an n value fixed at 1. The remainder of the ITC values can be found in Table 4.3.

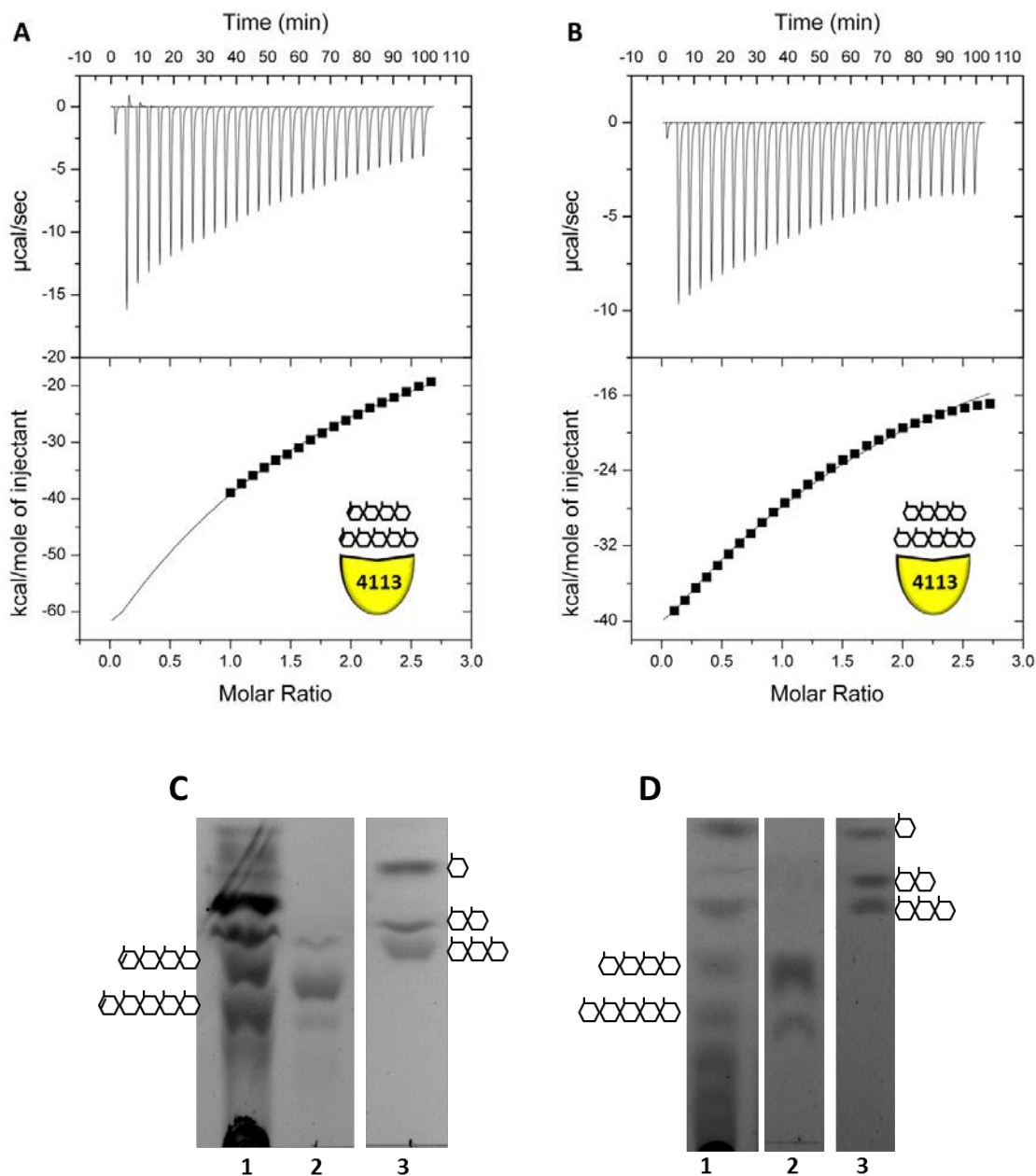


Figure 4.6: ITC binding isotherm of BT4113 (SusD-like) solute binding protein and (A) uGalA₅/uGalA₄ and (B) GalA₅/GalA₄. Top panel- Observed heat trace for each injection of the titration. Lower panel- Enthalpies of binding fit to a one site binding model. **(C)** TLC of uGalA₅/uGalA₄ product used for ITC analysis. Lane 1- BT4115-HG digestion products, 2- uGalA₅/uGalA₄ products post separation, 3- GalA, GalA₂ and GalA₃. **(D)** TLC of GalA₅/GalA₄ product used for ITC analysis. Lane 1- Polygalacturonase-HG digestion products, lane 2- GalA₅/GalA₄ products post separation, lane 3- GalA, GalA₂ and GalA₃. Product sizes are indicated by sizes in the left margin. Standard sizes are indicated by sizes in the right margin.

Table 4.3: ITC binding parameters obtained for BT4113 and BT4122

Protein	Ligand	<i>n</i>	K_a (10^3 M^{-1})	ΔH (kcal mol^{-1})	ΔS (kcal mol^{-1})	K_d (mM)
BT4113 SusD	GalA ₅ /GalA ₄	1	2.2 ± 0.2	-3.2 ± 0.2	1.4 ± 0.2	0.45
BT4113 SusD	uGalA ₅ /uGalA ₄	1	5.3 ± 1.3	-2.4 ± 0.3	2.6 ± 0.4	0.19
BT4122 SusD	GalA ₅ /GalA ₄	N.B.	N.B.	N.B.	N.B.	N.B.
BT4122 SusD	uGalA ₅ /uGalA ₄	N.B.	N.B.	N.B.	N.B.	N.B.

N.B.-No binding detected

4.3.4 BT4122

4.3.4.1 Cloning, protein expression and purification— The coding sequence of BT4122 cloned into the pET28a *E. coli* expression vector was obtained from our collaborators. BT4122 was expressed as an approximately 78 kDa soluble protein and purified by IMAC. BT4122 fractions were pooled based on SDS-PAGE analysis (Figure 4.3F) and dialyzed. Approximately 11 mg of BT4122 was obtained from 1 l culture as determined by A_{280} .

4.3.4.2 Affinity gel electrophoresis (AGE)— Based on its homology to SusD, BT4122 was screened for carbohydrate binding by AGE, as previously described. BT4122 displayed no significant shift in the presence of any of the screened polysaccharides (Figure 4.4).

4.3.4.3 ITC binding assays— Subsequent ITC analysis was attempted with HG and RG-I polysaccharides, and their constituent monosaccharides, GalA and Rha. BT4122 displayed no affinity to either the mono- or polysaccharides. Based on our understanding of the SusD role in the Sus system, we further attempted ITC analysis with both unsaturated and saturated oligogalacturonides obtained from large scale HG digestions (see sections 2.2.8 and 4.3.1.5). BT4122 displayed no detectable affinity for both saturated or unsaturated oligogalacturonides (Table 4.3).

4.4 Discussion

The recruitment of target starch polysaccharide to the cell surface of *B. thetaiotaomicron* has been previously reported in the *sus* system, with the SusCDEF proteins all having distinct roles¹³⁰. Without recruitment and subsequent transport of target polysaccharides into the periplasmic space, the complete hydrolysis of the target polysaccharide is halted, due to the compartmentalization of the proteins within different regions of the cell. Having demonstrated the systematic deconstruction of HG by enzymes from PUL75, we sought to focus our attention on the Sus-like transport machinery in an attempt to elucidate its role in PUL75 function. As previously described, PUL75 contains two copies of each of the SusC/SusD pair (Bt4114 and BT4121 and BT4113 and BT4122), and a pair of ORFs (BT4112 and BT4120). Though a general understanding of SusCDEF¹³⁰ roles in starch recruitment and transport exists, the roles of the homologous gene products coded in PUL75 remain undefined. In an attempt to elucidate these roles, and provide a rationale for the presence of multiple copies of Sus-like systems in PUL75, we sought to characterize the pair of ORF (*BT4112* and *BT4120*) and SusD-like (*BT4113* and *BT4122*) gene products.

Affinity Gel Electrophoresis (AGE) was performed on both SusD-like and ORF proteins (Figure 4.4). BT4113 and BT4122 did not interact with the select carbohydrates as there was no visible shift in migration when compared to the control. This would suggest that these homologs may interact with a different ligand or have specific requirements for the degree of oligomerization in the target ligand. On the other hand both BT4112 and BT4120, displayed a binding property for HG. BT4112 displayed a distinct shift in

migration in the presence of HG (Figure 4.4), indicating a strong interaction. BT4120, on the other hand, displayed a small shift in migration in the presence of HG, suggesting a weak interaction. These interactions provided the first evidence that these two proteins may in fact be ancillary binding proteins (SusE-like) and mediators of HG transport. In the AGE analysis a number of weak interactions were visualized (smearing or loss of protein in lane). These weak interactions could be the result of contaminated carbohydrate. As previously mentioned, it is difficult to extract and to completely purify pectic polysaccharides. Alternatively, they may very well just represent weak interactions between protein and ligand.

In order to quantify the interactions observed in AGE, ITC was used (Figure 4.5A). The weak interaction observed in the AGE gels between BT4120 and HG was not detected in ITC (Table 4.2). Conversely, the stronger interaction observed between BT4112 and HG generated heat and was quantified (Table 4.2). Based on our results, BT4112 acts as the initial recruiter of HG to the cell surface in PUL75, facilitating depolymerisation by the outer membrane PL1, BT4119. Present on the outer membrane surface, there BT4112 may localize into clusters. The formation of BT4112 clusters may create an avidity affect with regard to HG binding, further enhancing the previously described affinity between BT4112 and HG.

Further ITC analyses of the SusD-like proteins, BT4113 and BT4122, displayed no affinity for polymerized HG or RG-I. Subsequent ITC analyses of their substituent monosaccharides, GalA and Rha, also displayed no affinity. A previously characterized SusD displayed affinity for starch oligosaccharides⁷⁸, indicating that degree of

polymerization (DP) may have an effect on SusD affinity. In this regard, we generated GalA_n and uGalA_n (as described in 2.2.8 and 4.3.1.5) for further ITC analyses.

BT4122 displayed no affinity for GalA_n or uGalA_n, whereas BT4113 displayed affinity for both GalA_n and uGalA_n; where $n \geq 4$. The observed affinities are relatively weak compared to our susE-like affinities for polymerized HG (Table 4.2 & 4.3); however, given the role of SusD proteins in ligand transport, a strong affinity may impair effective transport of the oligosaccharides into the periplasmic space. Furthermore, SusD may mediate the formation of larger protein complexes due to the presence of TPRs observed in the crystal structure. In this light, the remaining members of the protein complex may increase the affinity *in vivo*, or they may interact with the SusC-like porin, BT4114.

Interestingly, no selectivity between GalA_n and uGalA_n was observed in BT4113 ITC analysis (Figure 4.6A-B and Table 4.3). This observation supports the endo-specific protein-ligand interaction previously described for malto oligosaccharides binding to SusD proteins⁸¹. At this time, GalA_n and uGalA_n with DP > 5 have yet to be purified for ITC analysis, and interactions with these ligands have yet to be tested.

There is an approximately 10 fold difference in the binding affinities between BT4112 and BT4113. The different functions of these binding proteins in HG recruitment (BT4112) and transport (BT4113) support these observations. Present on the cell surface, BT4112 is responsible for the initial recruitment and binding of polymerized HG. In the competitive environment of the distal gut, the stronger the affinity for a given carbohydrate, the less likely the carbohydrate will be sequestered by competing bacteria. On the other hand, BT4113, in concert with the other transport machinery (BT4114;

SusC-like) is responsible for the translocation of unsaturated oligogalacturonides into the periplasmic space. In this regard, a strong substrate affinity may impede downstream processing of unsaturated oligogalacturonides; however, an affinity ‘threshold’ must exist in order for the translocation to occur.

As previously mentioned, PUL75 contains two copies of the SusC-, (BT4114 and BT4121), SusD- (BT4113 and BT4122) and now identified susE-like proteins (BT4112 and BT4120). As demonstrated in Figure 1.7, two clusters of the Sus-like genes exist. Based on our results, we propose that HG is the target ligand for the cluster including BT4112, BT4113 and BT4114. More specifically, BT4112 is responsible for the recruitment of polymerized HG to the cell surface, and BT4113 is responsible for binding oligogalacturonides, facilitating their transport into the periplasmic space. Further experimentation to define the role of BT4114 in HG transport is required, with mutagenesis studies currently underway, to determine if an observed loss of function results when this mutant is grown on HG. Conversely, the function of the second cluster of Sus-like genes; BT4120, BT4121 and BT4122, remains unknown. Perhaps it may target a yet to be defined ligand for recruitment and transport.

4.5 Conclusion

The successful cloning and recombinant protein production of BT4112 and BT4120 facilitated their characterization and subsequent identification as *susE*-like proteins. BT4112 interacts with HG in AGE and ITC ($60 \pm 9 \times 10^3 \text{ M}^{-1}$). Similarly, the successful recombinant protein production of the two *SusD*-like proteins, BT4113 and BT4122, provided evidence of their differing roles in HG utilization. BT4113 displayed affinities for both GalA_n and uGalA_n, in ITC analysis, whereas BT4122 did not. BT4113 substrate recognition appears to be endo-specific, as there was no significant difference in GalA_n and uGalA_n affinities. Based on these analyses, we propose that the BT4112-14 cluster of *Sus*-like proteins are involved in HG utilization by facilitating the recruitment of polymerized HG to the cell surface (BT4112, *susE*-like) and facilitating the translocation of oligogalacturonides into the periplasm (BT4113 and BT4114, *SusC*- and *SusD*-like, respectively). The second cluster of *Sus*-like proteins (BT4120-22) displayed no affinity for HG, RG-I, PG or oligogalacturonides, and as such we hypothesize they target a yet to be identified carbohydrate. Most importantly, our characterization of BT4112 defined the recruitment of HG to the outer membrane, which would facilitate the outer membrane PL1 (BT4119) depolymerase activity at the cell surface. Similarly, the characterization of BT4113 defined the transport of uGalA_n products into the periplasm for further deconstruction by the PUL75 periplasmic enzymes; BT4115 and BT4116 (PLs), and BT4108 and BT4123 (GHs).

Chapter 5-

Conclusion and Future Directions

5.1 Overall Summary

Based on the data presented herein, we conclude that the gene products coded by PUL75 are responsible for the recruitment, transport and systematic deconstruction of polymerized HG by *B. thetaiotaomicron*. More specifically, we have elucidated the roles of the three PL1s (BT4115, BT4116 and BT4119) found in PUL75. Based on product profiles, and bioinformatics analyses, we propose that within *B. thetaiotaomicron*, the PL1s are compartmentalized. More specifically, we hypothesize that BT4119 is found on the outer membrane acting as a depolymerase, whereas BT4115 and BT4116 are found within the periplasm, producing shorter oligogalacturonides, providing a rationale for the multiple copies of PL1s present (Chapter 2) and consistent with other PUL systems. Subsequent characterization of the two GHs (BT4108 and BT4123) revealed activities responsible for further hydrolysis of PL1 generated uGalA_n products (Chapter 3). Most importantly, the discovery of BT4108's novel activity which removes the unsaturated moiety from PL1 generated uGalA_n, represents the key in overcoming a metabolic blockage for the complete hydrolysis of uGalA_n by BT4123. The identification and characterization of the susE-like protein (BT4112) provided a basis for the recruitment of polymerized HG to the cell surface, while the characterization of the SusD-like protein (BT4113) identified a role in the binding and proposed transport of oligogalacturonides into the periplasm (Chapter 4). Within PUL75, pairs of the SusC-, SusD- and SusE-like

proteins are present: however, the characterization of BT4112 and BT4113 demonstrates that these Sus-like proteins bind different ligands.

Based on our characterization of PUL75 gene products, we propose the following sequence of events to described HG recruitment, transport and utilization by *B. thetaiotaomicron*. A cartoon summary is presented in Figure 5.1.

Polymerized HG is recruited to the cell surface by BT4112 (Figure 5.1.i). Bound HG is cleaved into unsaturated oligosaccharides by the outer membrane bound PL1 BT4119 (Figure 5.1.ii). Oligogalacturonides are bound to BT4113, and subsequently translocated into the periplasm (Figure 5.1.iii). Translocated oligogalacturonides are further deconstructed by the remaining periplasmic PLs (BT4115 and BT4116) (Figure 5.1.iv). BT4108 removes the unsaturated moiety from the non-reducing end, resulting in the production of linear DKI and saturated oligosaccharides (Figure 5.1.v). BT4123 hydrolyses the remaining saturated oligogalacturonides into monosaccharides and disaccharides (GalA and GalA₂, respectively) (Figure 5.1.vi). DKI and GalA are translocated into the cytoplasm by a yet to be identified transport machinery (Figure 5.1.vii).

Partial reconstitution of the PUL75 pathway *in vitro* demonstrated the generation of GalA and GalA₂ products from polymerized HG (Figure 3.10). This is consistent with the analysis of supernatant from wild type *B. thetaiotaomicron* growth on HG (Figure 5.2), which also displays the presence of small oligosaccharides. These results provide biological evidence supporting our *in vitro* reconstituted pathway.

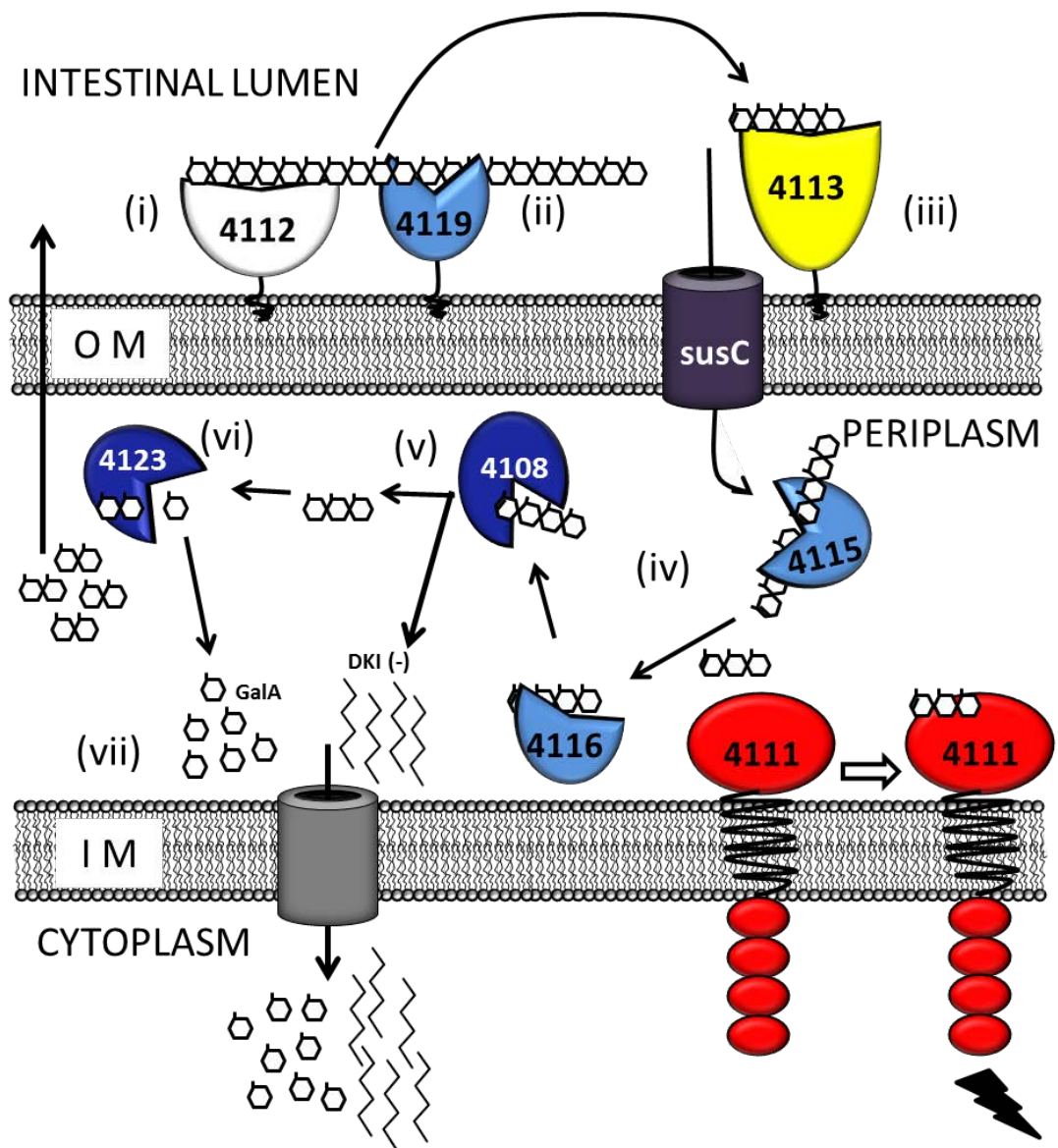


Figure 5.1: Proposed model of PUL75 systematic deconstruction of HG. HG is recruited from the intestine by BT4112 (i). BT4119 depolymerizes HG into oligogalacturonides (ii). Oligogalacturonides are bound by BT4113 and transported into the periplasm through the yet to be characterized SusC-like porin (iii). BT4115 and BT4116 further hydrolyze oligogalacturonides (iv). BT4108 removes the unsaturated moiety from the oligogalacturonides resulting in the production of linear DKI (v). BT4123 cleaves the remaining saturated oligogalacturonides into monosaccharides and disaccharides (vi). Monosaccharides and DKI are transported into the cytoplasm (vii). BT4111 are activated by an oligogalacturonide of a length yet to be determined. (The uGalA₃ is for illustrative purposes only). (Appendix 5) OM- Outer membrane, IM-Inner membrane

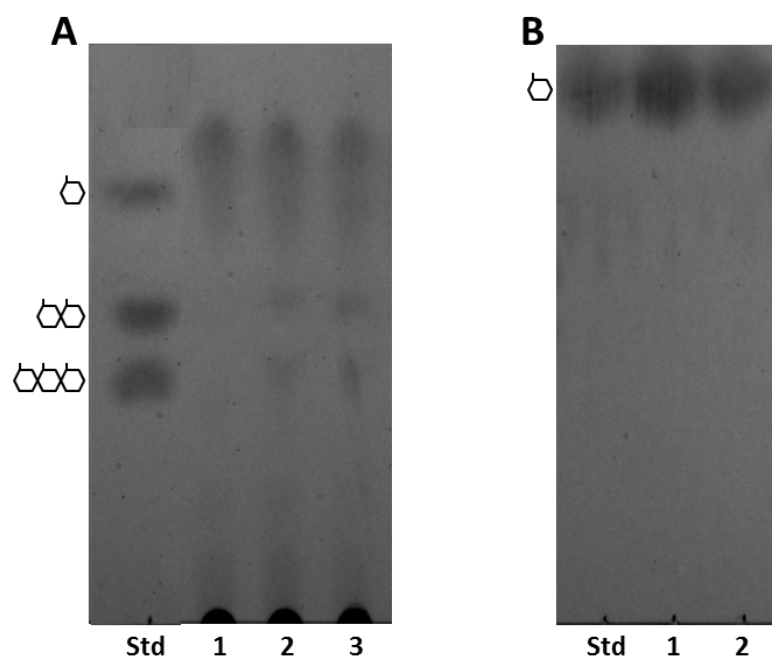


Figure 5.2: TLC analysis of supernatant from wild type *B. thetaiotaomicron* grown on (A) HG and (B) glucose. (A) Std- GalA, GalA₂ and GalA₃. Lane 1- Supernatant prior to inoculation, lane 2- 12 hr post inoculation (replicate 1) and lane 3- 12 hr post inoculation (replicate 2). **(B)** Std- glucose. Lane 1- supernatant prior to inoculation and lane 2- 12 hr post inoculation. All samples were grown in identical anaerobic conditions with supernatant aliquots taken at the indicated time points.

5.2 Future Directions

Based on the data presented to date a complete understanding of the defined roles and function of the gene products coded by PUL75 of *B. thetaiotaomicron* has been realized; however, there still remains a large gap in our structural understanding of these proteins. Elucidating the molecular mechanism is a priority. Structure studies can help and progress has been made in these areas.

In order to define the target polysaccharide and provide a rationale for the multiple copies of PUL75 gene products, further analysis and characterization of PUL75's second group of SusC-, SusD- and susE-like proteins (BT4120-22) and corresponding HTCS (BT4124) is required.

Further efforts to create selective mutations, targeting specific genes within PUL75, would allow for new *B. thetaiotaomicron* phenotypes to be observed and identify essential genes for PUL75 HG utilization. Of particular interest would be a $\Delta BT4119$ mutant to further investigate the requirement for a cell surface depolymerase and a $\Delta BT4108$ mutant to investigate the unique BT4108 activity as described in chapter 3.

References Cited

1. Rolland, F., Baena-Gonzalez, E., and Sheen, J. (2006) Sugar sensing and signaling in plants: conserved and novel mechanisms, *Annual review of plant biology* 57, 675-709.
2. Wrzosek, L., Miquel, S., Noordine, M. L., Bouet, S., Joncquel Chevalier-Curt, M., Robert, V., Philippe, C., Bridonneau, C., Cherbuy, C., Robbe-Masselot, C., Langella, P., and Thomas, M. (2013) *Bacteroides thetaiotaomicron* and *Faecalibacterium prausnitzii* influence the production of mucus glycans and the development of goblet cells in the colonic epithelium of a gnotobiotic model rodent, *BMC biology* 11, 61.
3. Voet, D., Voet, J. G., and Pratt, C. W. (2008) *Fundamentals of biochemistry*, Third ed., John Wiley & Sons, New Jersey.
4. Burton, R. A., Gidley, M. J., and Fincher, G. B. (2010) Heterogeneity in the chemistry, structure and function of plant cell walls, *Nature chemical biology* 6, 724-732.
5. Gilbert, H. J. (2010) The biochemistry and structural biology of plant cell wall deconstruction, *Plant physiology* 153, 444-455.
6. Mohnen, D., Bar-Peled, M., and Somerville, C. (2009) Cell wall polysaccharide synthesis, In *Biomass Recalcitrance*, pp 94-187, Blackwell Publishing Ltd.
7. Jarvis, M. C., Briggs, S. P. H., and Knox, J. P. (2003) Intercellular adhesion and cell separation in plants, *Plant, Cell and Environment* 26, 977-989.
8. Popper, Z. A., Michel, G., Herve, C., Domozych, D. S., Willats, W. G., Tuohy, M. G., Kloareg, B., and Stengel, D. B. (2011) Evolution and diversity of plant cell walls: from algae to flowering plants, *Annual review of plant biology* 62, 567-590.
9. Domozych, D. S. (2001) Algal cell walls, In *eLS*, John Wiley & Sons, Ltd.
10. Selvendran, R. R., and O'Neill, M. A. (1987) Isolation and analysis of cell walls from plant material, *Methods of biochemical analysis* 32, 25-153.
11. Thakur, B. R., Singh, R. K., and Handa, A. K. (1997) Chemistry and uses of pectin--a review, *Critical reviews in food science and nutrition* 37, 47-73.
12. Atmodjo, M. A., Hao, Z., and Mohnen, D. (2013) Evolving views of pectin biosynthesis, *Annual Review of Plant Biology* 64, 747-779.
13. Scavetta, R. D., Herron, S. R., Hotchkiss, A. T., Kita, N., Keen, N. T., Benen, J. A., Kester, H. C., Visser, J., and Jurnak, F. (1999) Structure of a plant cell wall fragment complexed to pectate lyase C, *The Plant cell* 11, 1081-1092.

14. Jensen, M. H., Otten, H., Christensen, U., Borchert, T. V., Christensen, L. L., Larsen, S., and Leggio, L. L. (2010) Structural and biochemical studies elucidate the mechanism of rhamnogalacturonan lyase from *Aspergillus aculeatus*, *J Mol Biol* 404, 100-111.
15. Fries, M., Ihrig, J., Brocklehurst, K., Shevchik, V. E., and Pickersgill, R. W. (2007) Molecular basis of the activity of the phytopathogen pectin methylesterase, *Embo J* 26, 3879-3887.
16. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot, *Acta Crystallogr D Biol Crystallogr* 66, 486-501.
17. O'Neill, M. A., Warrenfeltz, D., Kates, K., Pellerin, P., Doco, T., Darvill, A. G., and Albersheim, P. (1996) Rhamnogalacturonan-II, a pectic polysaccharide in the walls of growing plant cell, forms a dimer that is covalently cross-linked by a borate ester. In vitro conditions for the formation and hydrolysis of the dimer, *The Journal of biological chemistry* 271, 22923-22930.
18. Liners, F., Letesson, J. J., Didembourg, C., and Van Cutsem, P. (1989) Monoclonal antibodies against pectin: Recognition of a conformation induced by calcium, *Plant physiology* 91, 1419-1424.
19. O'Neill, M. A., Ishii, T., Albersheim, P., and Darvill, A. G. (2004) Rhamnogalacturonan II: structure and function of a borate cross-linked cell wall pectic polysaccharide, *Annual review of plant biology* 55, 109-139.
20. Caffall, K. H., and Mohnen, D. (2009) The structure, function, and biosynthesis of plant cell wall pectic polysaccharides, *Carbohydrate research* 344, 1879-1900.
21. Bar-Peled, M., Urbanowicz, B. R., and O'Neill, M. A. (2012) The synthesis and origin of the pectic polysaccharide rhamnogalacturonan II - Insights from nucleotide sugar formation and diversity, *Frontiers in plant science* 3, 92.
22. Willats, W. G., McCartney, L., Mackie, W., and Knox, J. P. (2001) Pectin: cell biology and prospects for functional analysis, *Plant molecular biology* 47, 9-27.
23. Brummell, D. A. (2006) Cell wall disassembly in ripening fruit, *Functional Plant Biology* 33, 103.
24. Ridley, B. L., O'Neill, M. A., and Mohnen, D. (2001) Pectins: structure, biosynthesis, and oligogalacturonide-related signaling, *Phytochemistry* 57, 929-967.
25. Savatin, D. V., Ferrari, S., Sicilia, F., and De Lorenzo, G. (2011) Oligogalacturonide-auxin antagonism does not require posttranscriptional gene silencing or stabilization of auxin response repressors in *Arabidopsis*, *Plant physiology* 157, 1163-1174.

26. Ferrari, S., Savatin, D. V., Sicilia, F., Gramegna, G., Cervone, F., and Lorenzo, G. D. (2013) Oligogalacturonides: plant damage-associated molecular patterns and regulators of growth and development, *Frontiers in plant science* 4, 49.
27. Sterling, J. D., Atmodjo, M. A., Inwood, S. E., Kumar Kolli, V. S., Quigley, H. F., Hahn, M. G., and Mohnen, D. (2006) Functional identification of an *Arabidopsis* pectin biosynthetic homogalacturonan galacturonosyltransferase, *Proceedings of the National Academy of Sciences of the United States of America* 103, 5236-5241.
28. Cid, M., Pedersen, H. L., Kaneko, S., Coutinho, P. M., Henrissat, B., Willats, W. G., and Boraston, A. B. (2010) Recognition of the helical structure of beta-1,4-galactan by a new family of carbohydrate-binding modules, *The Journal of biological chemistry* 285, 35999-36009.
29. Mohnen, D. (2008) Pectin structure and biosynthesis, *Current opinion in plant biology* 11, 266-277.
30. Willats, W. G., McCartney, L., and Knox, J. P. (2001) In-situ analysis of pectic polysaccharides in seed mucilage and at the root surface of *Arabidopsis thaliana*, *Planta* 213, 37-44.
31. Jarvis, M. C., and Apperley, D. C. (1995) Chain conformation in concentrated pectic gels: evidence from ¹³C NMR, *Carbohydrate research* 275, 131-145.
32. Nakamura, A., Furuta, H., Maeda, H., Takao, T., and Nagamatsu, Y. (2002) Structural studies by stepwise enzymatic degradation of the main backbone of soybean soluble polysaccharides consisting of galacturonan and rhamnogalacturonan, *Bioscience, biotechnology, and biochemistry* 66, 1301-1313.
33. Lau, J. M., McNeil, M., Darvill, A. G., and Albersheim, P. (1985) Structure of the backbone of rhamnogalacturonan I, a pectic polysaccharide in the primary cell walls of plants, *Carbohydrate research* 137, 111-125.
34. Atmodjo, M. A., Hao, Z., and Mohnen, D. (2013) Evolving views of pectin biosynthesis, *Annu Rev Plant Biol* 64, 747-779.
35. J, O. B., Harholt, J., Scheller, H. V., and Orfila, C. (2004) Rhamnogalacturonan I in *Solanum tuberosum* tubers contains complex arabinogalactan structures, *Phytochemistry* 65, 1429-1438.
36. Freshour, G., Clay, R. P., Fuller, M. S., Albersheim, P., Darvill, A. G., and Hahn, M. G. (1996) Developmental and tissue-specific structural alterations of the cell-wall polysaccharides of *Arabidopsis thaliana* roots, *Plant physiology* 110, 1413-1429.

37. Guillemin, F., Guillon, F., Bonnin, E., Devaux, M. F., Chevalier, T., Knox, J. P., Liners, F., and Thibault, J. F. (2005) Distribution of pectic epitopes in cell walls of the sugar beet root, *Planta* 222, 355-371.
38. Thomas, J. R., McNeil, M., Darvill, A. G., and Albersheim, P. (1987) Structure of plant cell walls : XIX. Isolation and characterization of wall polysaccharides from suspension-cultured Douglas fir cells, *Plant physiology* 83, 659-671.
39. Spellman, M. W., McNeil, M., Darvill, A. G., Albersheim, P., and Dell, A. (1983) Characterization of a structurally complex heptasaccharide isolated from the pectic polysaccharide rhamnogalacturonan II, *Carbohydrate research* 122, 131-153.
40. Stevenson, T. T., Darvill, A. G., and Albersheim, P. (1988) 3-deoxy-d-lyxo-2-heptulosaric acid, a component of the plant cell-wall polysaccharide rhamnogalacturonan-II, *Carbohydrate research* 179, 269-288.
41. York, W. S., Darvill, A. G., McNeil, M., and Albersheim, P. (1985) 3-deoxy-d-manno-2-octulosonic acid (KDO) is a component of rhamnogalacturonan II, a pectic polysaccharide in the primary cell walls of plants, *Carbohydrate research* 138, 109-126.
42. O'Neill, M. A., Eberhard, S., Albersheim, P., and Darvill, A. G. (2001) Requirement of borate cross-linking of cell wall rhamnogalacturonan II for *Arabidopsis* growth, *Science* 294, 846-849.
43. Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics, *Nucleic acids research* 37, D233-238.
44. Brummell, D. A., and Harpster, M. H. (2001) Cell wall metabolism in fruit softening and quality and its manipulation in transgenic plants, *Plant molecular biology* 47, 311-340.
45. Toth, I. K., Bell, K. S., Holeva, M. C., and Birch, P. R. (2003) Soft rot *Erwiniae*: from genes to genomes, *Molecular plant pathology* 4, 17-30.
46. Giovane, A., Servillo, L., Balestrieri, C., Raiola, A., D'Avino, R., Tamburrini, M., Ciardiello, M. A., and Camardella, L. (2004) Pectin methylesterase inhibitor, *Biochimica et biophysica acta* 1696, 245-252.
47. Pelloux, J., Rusterucci, C., and Mellerowicz, E. J. (2007) New insights into pectin methylesterase structure and function, *Trends in plant science* 12, 267-277.
48. Di Matteo, A., Giovane, A., Raiola, A., Camardella, L., Bonivento, D., De Lorenzo, G., Cervone, F., Bellincampi, D., and Tsernoglou, D. (2005) Structural

basis for the interaction between pectin methylesterase and a specific inhibitor protein, *The Plant cell* 17, 849-858.

49. Bellincampi, D., Dipierro, N., Salvi, G., Cervone, F., and De Lorenzo, G. (2000) Extracellular H₂O₂ induced by oligogalacturonides is not involved in the inhibition of the auxin-regulated rolB gene expression in tobacco leaf explants, *Plant physiology* 122, 1379-1385.
50. Davis, K. R., and Hahlbrock, K. (1987) Induction of defense responses in cultured parsley cells by plant cell wall fragments, *Plant physiology* 84, 1286-1290.
51. Broekaert, W. F., Lambrechts, D., Verbelen, J. P., and Peumans, W. J. (1988) *Datura stramonium* agglutinin : Location in the seed and release upon imbibition, *Plant physiology* 86, 569-574.
52. Herron, S. R., Benen, J. A., Scavetta, R. D., Visser, J., and Jurnak, F. (2000) Structure and function of pectic enzymes: virulence factors of plant pathogens, *Proc Natl Acad Sci U S A* 97, 8762-8769.
53. Abbott, D. W., and Boraston, A. B. (2008) Structural biology of pectin degradation by Enterobacteriaceae, *Microbiol Mol Biol Rev* 72, 301-316, table of contents.
54. Yip, V. L., and Withers, S. G. (2006) Breakdown of oligosaccharides by the process of elimination, *Current opinion in chemical biology* 10, 147-155.
55. Naumoff, D. G. (2011) Hierarchical classification of glycoside hydrolases, *Biochemistry. Biokhimiia* 76, 622-635.
56. Davies, G. J., Wilson, K. S., and Henrissat, B. (1997) Nomenclature for sugar-binding subsites in glycosyl hydrolases, *The Biochemical journal* 321 (Pt 2), 557-559.
57. Garron, M. L., and Cygler, M. (2010) Structural and mechanistic classification of uronic acid-containing polysaccharide lyases, *Glycobiology* 20, 1547-1573.
58. Lombard, V., Bernard, T., Rancurel, C., Brumer, H., Coutinho, P. M., and Henrissat, B. (2010) A hierarchical classification of polysaccharide lyases for glycogenomics, *Biochem J* 432, 437-444.
59. James, V., and Hugouvieux-Cotte-Pattat, N. (1996) Regulatory systems modulating the transcription of the pectinase genes of *Erwinia chrysanthemi* are conserved in *Escherichia coli*, *Microbiology* 142 (Pt 9), 2613-2619.
60. Lara-Marquez, A., Zavala-Paramo, M. G., Lopez-Romero, E., and Camacho, H. C. (2011) Biotechnological potential of pectinolytic complexes of fungi, *Biotechnol Lett* 33, 859-868.

61. Kikot, G. E., Hours, R. A., and Alconada, T. M. (2009) Contribution of cell wall degrading enzymes to pathogenesis of *Fusarium graminearum*: a review, *J Basic Microbiol* 49, 231-241.
62. <http://commonfund.nih.gov/hmp/index>. Human Microbiome Project.
63. Koropatkin, N. M., Cameron, E. A., and Martens, E. C. (2012) How glycan metabolism shapes the human gut microbiota, *Nature reviews. Microbiology* 10, 323-335.
64. McNeil, N. I. (1984) The contribution of the large intestine to energy supplies in man, *The American journal of clinical nutrition* 39, 338-342.
65. Moore, W. E., Cato, E. P., and Holdeman, L. V. (1978) Some current concepts in intestinal bacteriology, *The American journal of clinical nutrition* 31, S33-42.
66. Lindberg, A. A., Weintraub, A., Zahringer, U., and Rietschel, E. T. (1990) Structure-activity relationships in lipopolysaccharides of *Bacteroides fragilis*, *Reviews of infectious diseases* 12 Suppl 2, S133-141.
67. Strohm, H., Payne, C. M., and Ryan, K. J. (1983) Demonstration of *Bacteroides* capsules by light microscopy and ultrastructural cytochemistry, *American journal of clinical pathology* 79, 591-597.
68. Wexler, H. M. (2007) *Bacteroides*: the good, the bad, and the nitty-gritty, *Clinical microbiology reviews* 20, 593-621.
69. Moore, W. E., and Holdeman, L. V. (1974) Human fecal flora: the normal flora of 20 Japanese-Hawaiians, *Applied microbiology* 27, 961-979.
70. Xu, J., Bjursell, M. K., Himrod, J., Deng, S., Carmichael, L. K., Chiang, H. C., Hooper, L. V., and Gordon, J. I. (2003) A genomic view of the human-*Bacteroides thetaiotaomicron* symbiosis, *Science* 299, 2074-2076.
71. Kuwahara, T., Yamashita, A., Hirakawa, H., Nakayama, H., Toh, H., Okada, N., Kuhara, S., Hattori, M., Hayashi, T., and Ohnishi, Y. (2004) Genomic analysis of *Bacteroides fragilis* reveals extensive DNA inversions regulating cell surface adaptation, *Proc Natl Acad Sci U S A* 101, 14919-14924.
72. Cerdeno-Tarraga, A. M., Patrick, S., Crossman, L. C., Blakely, G., Abratt, V., Lennard, N., Poxton, I., Duerden, B., Harris, B., Quail, M. A., Barron, A., Clark, L., Corton, C., Doggett, J., Holden, M. T., Larke, N., Line, A., Lord, A., Norbertczak, H., Ormond, D., Price, C., Rabinowitsch, E., Woodward, J., Barrell, B., and Parkhill, J. (2005) Extensive DNA inversions in the *B. fragilis* genome control variable gene expression, *Science* 307, 1463-1465.

73. Krinos, C. M., Coyne, M. J., Weinacht, K. G., Tzianabos, A. O., Kasper, D. L., and Comstock, L. E. (2001) Extensive surface diversity of a commensal microorganism by multiple DNA inversions, *Nature* 414, 555-558.
74. Bolam, D. N., and Sonnenburg, J. L. (2011) Mechanistic insight into polysaccharide use within the intestinal microbiota, *Gut microbes* 2, 86-90.
75. Martens, E. C., Lowe, E. C., Chiang, H., Pudlo, N. A., Wu, M., McNulty, N. P., Abbott, D. W., Henrissat, B., Gilbert, H. J., Bolam, D. N., and Gordon, J. I. (2011) Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts, *PLoS biology* 9, e1001221.
76. Sonnenburg, E. D., Sonnenburg, J. L., Manchester, J. K., Hansen, E. E., Chiang, H. C., and Gordon, J. I. (2006) A hybrid two-component system protein of a prominent human gut symbiont couples glycan sensing in vivo to carbohydrate metabolism, *Proceedings of the National Academy of Sciences of the United States of America* 103, 8834-8839.
77. Reeves, A. R., Wang, G. R., and Salyers, A. A. (1997) Characterization of four outer membrane proteins that play a role in utilization of starch by *Bacteroides thetaiotaomicron*, *J Bacteriol* 179, 643-649.
78. Koropatkin, N., Martens, E. C., Gordon, J. I., and Smith, T. J. (2009) Structure of a SusD homologue, BT1043, involved in mucin O-glycan utilization in a prominent human gut symbiont, *Biochemistry* 48, 1532-1542.
79. Reeves, A. R., D'Elia, J. N., Frias, J., and Salyers, A. A. (1996) A *Bacteroides thetaiotaomicron* outer membrane protein that is essential for utilization of maltooligosaccharides and starch, *J Bacteriol* 178, 823-830.
80. Shipman, J. A., Berleman, J. E., and Salyers, A. A. (2000) Characterization of four outer membrane proteins involved in binding starch to the cell surface of *Bacteroides thetaiotaomicron*, *J Bacteriol* 182, 5365-5372.
81. Koropatkin, N. M., Martens, E. C., Gordon, J. I., and Smith, T. J. (2008) Starch catabolism by a prominent human gut symbiont is directed by the recognition of amylose helices, *Structure* 16, 1105-1115.
82. Sonnenburg, E. D., Zheng, H., Joglekar, P., Higginbottom, S. K., Firbank, S. J., Bolam, D. N., and Sonnenburg, J. L. (2010) Specificity of polysaccharide use in intestinal *Bacteroides* species determines diet-induced microbiota alterations, *Cell* 141, 1241-1252.
83. Cameron, E. A., Maynard, M. A., Smith, C. J., Smith, T. J., Koropatkin, N. M., and Martens, E. C. (2012) Multidomain carbohydrate-binding proteins involved in *Bacteroides thetaiotaomicron* starch metabolism, *The Journal of biological chemistry* 287, 34614-34625.

84. Reid, G. (2004) When microbe meets human, *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 39, 827-830.
85. Marcobal, A., Barboza, M., Sonnenburg, E. D., Pudlo, N., Martens, E. C., Desai, P., Lebrilla, C. B., Weimer, B. C., Mills, D. A., German, J. B., and Sonnenburg, J. L. (2011) *Bacteroides* in the infant gut consume milk oligosaccharides via mucus-utilization pathways, *Cell host & microbe* 10, 507-514.
86. Salyers, A. A., Bonheyo, G., and Shoemaker, N. B. (2000) Starting a new genetic system: lessons from *Bacteroides*, *Methods* 20, 35-46.
87. Charnock, S. J., Brown, I. E., Turkenburg, J. P., Black, G. W., and Davies, G. J. (2002) Convergent evolution sheds light on the anti-beta -elimination mechanism common to family 1 and 10 polysaccharide lyases, *Proceedings of the National Academy of Sciences of the United States of America* 99, 12067-12072.
88. Abbott, D. W., and Boraston, A. B. (2007) A family 2 pectate lyase displays a rare fold and transition metal-assisted beta-elimination, *J Biol Chem* 282, 35328-35336.
89. Abbott, D. W., Gilbert, H. J., and Boraston, A. B. (2010) The active site of oligogalacturonate lyase provides unique insights into cytoplasmic oligogalacturonate beta-elimination, *J Biol Chem* 285, 39029-39038.
90. Tardy, F., Nasser, W., Robert-Baudouy, J., and Hugouvieux-Cotte-Pattat, N. (1997) Comparative analysis of the five major *Erwinia chrysanthemi* pectate lyases: enzyme characteristics and potential inhibitors, *J Bacteriol* 179, 2503-2511.
91. Abbott, D. W., Thomas, D., Pluvinaige, B., and Boraston, A. B. (2013) An ancestral member of the polysaccharide lyase family 2 displays endolytic activity and magnesium dependence, *Appl Biochem Biotechnol* 171, 1911-1923.
92. Seyedarabi, A., To, T. T., Ali, S., Hussain, S., Fries, M., Madsen, R., Clausen, M. H., Teixeira, S., Brocklehurst, K., and Pickersgill, R. W. (2010) Structural insights into substrate specificity and the anti beta-elimination mechanism of pectate lyase, *Biochemistry* 49, 539-546.
93. Yoder, M. D., Keen, N. T., and Jurnak, F. (1993) New domain motif: the structure of pectate lyase C, a secreted plant virulence factor, *Science* 260, 1503-1507.
94. Jenkins, J., and Pickersgill, R. (2001) The architecture of parallel beta-helices and related folds, *Prog Biophys Mol Biol* 77, 111-175.
95. Jenkins, J., Mayans, O., and Pickersgill, R. (1998) Structure and evolution of parallel beta-helix proteins, *J Struct Biol* 122, 236-246.

96. Jenkins, J., Shevchik, V. E., Hugouvieux-Cotte-Pattat, N., and Pickersgill, R. W. (2004) The crystal structure of pectate lyase Pel9A from *Erwinia chrysanthemi*, *J Biol Chem* 279, 9139-9145.
97. Wilkins, M. R., Gasteiger, E., Bairoch, A., Sanchez, J. C., Williams, K. L., Appel, R. D., and Hochstrasser, D. F. (1999) Protein identification and analysis tools in the ExPASy server, *Methods Mol Biol* 112, 531-552.
98. Xiao, Z., Bergeron, H., Grosse, S., Beauchemin, M., Garron, M. L., Shaya, D., Sulea, T., Cygler, M., and Lau, P. C. (2008) Improvement of the thermostability and activity of a pectate lyase by single amino acid substitutions, using a strategy based on melting-temperature-guided sequence alignment, *Applied and environmental microbiology* 74, 1183-1189.
99. Guillen Schlippe, Y. V., and Hedstrom, L. (2005) Guanidine derivatives rescue the Arg418Ala mutation of *Trichomonas foetus* IMP dehydrogenase, *Biochemistry* 44, 16695-16700.
100. Biely, P., Benen, J., Heinrichova, K., Kester, H. C., and Visser, J. (1996) Inversion of configuration during hydrolysis of alpha-1,4-galacturonidic linkage by three *Aspergillus* polygalacturonases, *FEBS Lett* 382, 249-255.
101. Petersen, T. N., Kauppinen, S., and Larsen, S. (1997) The crystal structure of rhamnogalacturonase A from *Aspergillus aculeatus*: a right-handed parallel beta helix, *Structure* 5, 533-544.
102. Pickersgill, R., Smith, D., Worboys, K., and Jenkins, J. (1998) Crystal structure of polygalacturonase from *Erwinia carotovora* ssp. *carotovora*, *J Biol Chem* 273, 24660-24664.
103. Abbott, D. W., and Boraston, A. B. (2007) The structural basis for exopolygalacturonase activity in a family 28 glycoside hydrolase, *J Mol Biol* 368, 1215-1222.
104. Shimizu, T., Nakatsu, T., Miyairi, K., Okuno, T., and Kato, H. (2002) Active-site architecture of endopolygalacturonase I from *Stereum purpureum* revealed by crystal structures in native and ligand-bound forms at atomic resolution, *Biochemistry* 41, 6651-6659.
105. Vasella, A., Davies, G. J., and Bohm, M. (2002) Glycosidase mechanisms, *Curr Opin Chem Biol* 6, 619-629.
106. Itoh, T., Ochiai, A., Mikami, B., Hashimoto, W., and Murata, K. (2006) A novel glycoside hydrolase family 105: the structure of family 105 unsaturated rhamnogalacturonyl hydrolase complexed with a disaccharide in comparison with family 88 enzyme complexed with the disaccharide, *J Mol Biol* 360, 573-585.

107. Itoh, T., Ochiai, A., Mikami, B., Hashimoto, W., and Murata, K. (2006) Structure of unsaturated rhamnogalacturonyl hydrolase complexed with substrate, *Biochemical and biophysical research communications* 347, 1021-1029.
108. Mill, P. J., and Tuttobello, R. (1961) The pectic enzymes of *Aspergillus niger* *The Biochemical journal* 79, 57-64.
109. Hasegawa, S., and Nagel, C. W. (1968) Isolation of an oligogalacturonate hydrolase from a *Bacillus* species, *Archives of biochemistry and biophysics* 124, 513-520.
110. Itoh, T., Ochiai, A., Mikami, B., Hashimoto, W., and Murata, K. (2006) Structure of unsaturated rhamnogalacturonyl hydrolase complexed with substrate, *Biochem Biophys Res Commun* 347, 1021-1029.
111. Kofod, L. V., Kauppinen, S., Christgau, S., Andersen, L. N., Heldt-Hansen, H. P., Dorreich, K., and Dalboge, H. (1994) Cloning and characterization of two structurally and functionally divergent rhamnogalacturonases from *Aspergillus aculeatus*, *The Journal of biological chemistry* 269, 29182-29189.
112. Mutter, M., Beldman, G., Pitson, S. M., Schols, H. A., and Voragen, A. G. (1998) Rhamnogalacturonan alpha-d-galactopyranosyluronohydrolase. An enzyme that specifically removes the terminal nonreducing galacturonosyl residue in rhamnogalacturonan regions of pectin, *Plant Physiol* 117, 153-163.
113. Pitson, S. M., Mutter, M., van den Broek, L. A., Voragen, A. G., and Beldman, G. (1998) Stereochemical course of hydrolysis catalysed by alpha-L-rhamnosyl and alpha-D-galacturonosyl hydrolases from *Aspergillus aculeatus*, *Biochemical and biophysical research communications* 242, 552-559.
114. Khan, M., Nakkeeran, E., and Umesh-Kumar, S. (2013) Potential application of pectinase in developing functional foods, *Annual review of food science and technology* 4, 21-34.
115. Lang, C., and Dornenburg, H. (2000) Perspectives in the biological function and the technological application of polygalacturonases, *Applied microbiology and biotechnology* 53, 366-375.
116. Nyvall-Collen, P., Jeudy, A., Sassi, J. F., Groisillier, A., Czjzek, M., Coutinho, P. M., and Helbert, W. (2014) A novel unsaturated beta-glucuronyl hydrolase involved in ulvan degradation unveils the versatility of stereochemistry requirements in family GH105, *J Biol Chem*.
117. Pijning, T., van Pouderoyen, G., Kluskens, L., van der Oost, J., and Dijkstra, B. W. (2009) The crystal structure of a hyperthermoactive exopolygalacturonase from *Thermotoga maritima* reveals a unique tetramer, *FEBS Lett* 583, 3665-3670.

118. Zhang, R., Minh, T., Lezondra, L., Korolev, S., Moy, S. F., Collart, F., and Joachimiak, A. (2005) 1.6 Å crystal structure of YteR protein from *Bacillus subtilis*, a predicted lyase, *Proteins* 60, 561-565.
119. Wilkins, M. R., Gasteiger, E., Bairoch, A., Sanchez, J. C., Williams, K. L., Appel, R. D., and Hochstrasser, D. F. (1999) Protein identification and analysis tools in the ExPASy server, *Methods Mol Biol* 112, 531-552.
120. Miller, G. L. (1959) Use of dinitrosalicylic acid reagent for determination of reducing sugar, *Analytical Chemistry* 31, 426-428.
121. Kluskens, L. D., van Alebeek, G. J., Walther, J., Voragen, A. G., de Vos, W. M., and van der Oost, J. (2005) Characterization and mode of action of an exopolygalacturonase from the hyperthermophilic bacterium *Thermotoga maritima*, *The FEBS journal* 272, 5464-5473.
122. Harris, T. K., and Turner, G. J. (2002) Structural basis of perturbed pK_a values of catalytic groups in enzyme active sites, *IUBMB life* 53, 85-98.
123. Blatch, G. L., and Lassle, M. (1999) The tetratricopeptide repeat: a structural motif mediating protein-protein interactions, *BioEssays : news and reviews in molecular, cellular and developmental biology* 21, 932-939.
124. Abbott, D. W., and van Bueren, A. L. (2014) Using structure to inform carbohydrate binding module function, *Current opinion in structural biology* 28C, 32-40.
125. Wimmerova, M., Kozmon, S., Necasova, I., Mishra, S. K., Komarek, J., and Koca, J. (2012) Stacking interactions between carbohydrate and protein quantified by combination of theoretical and experimental methods, *PloS one* 7, e46032.
126. Abbott, D. W., Hrynuik, S., and Boraston, A. B. (2007) Identification and characterization of a novel periplasmic polygalacturonic acid binding protein from *Yersinia enterocolitica*, *J Mol Biol* 367, 1023-1033.
127. Abbott, D. W., and Boraston, A. B. (2012) Quantitative approaches to the analysis of carbohydrate-binding module function, *Methods Enzymol* 510, 211-231.
128. MicroCal, L. VP-ITC microcalorimeter user's manual.
129. Cameron, R. G., Luzio, G. A., Kim, Y., Vasu, P., Savary, B. J., and Williams, M. A. K. (2011) Characterization of nanostructural modifications introduced into a model pectic homogalacturonan by esterases or chemical saponification and modeling of enzyme mode of action, In *Nanotechnology 2011: Bio Sensors, Instruments, Medical, Environment and Energy*, pp 275-278.

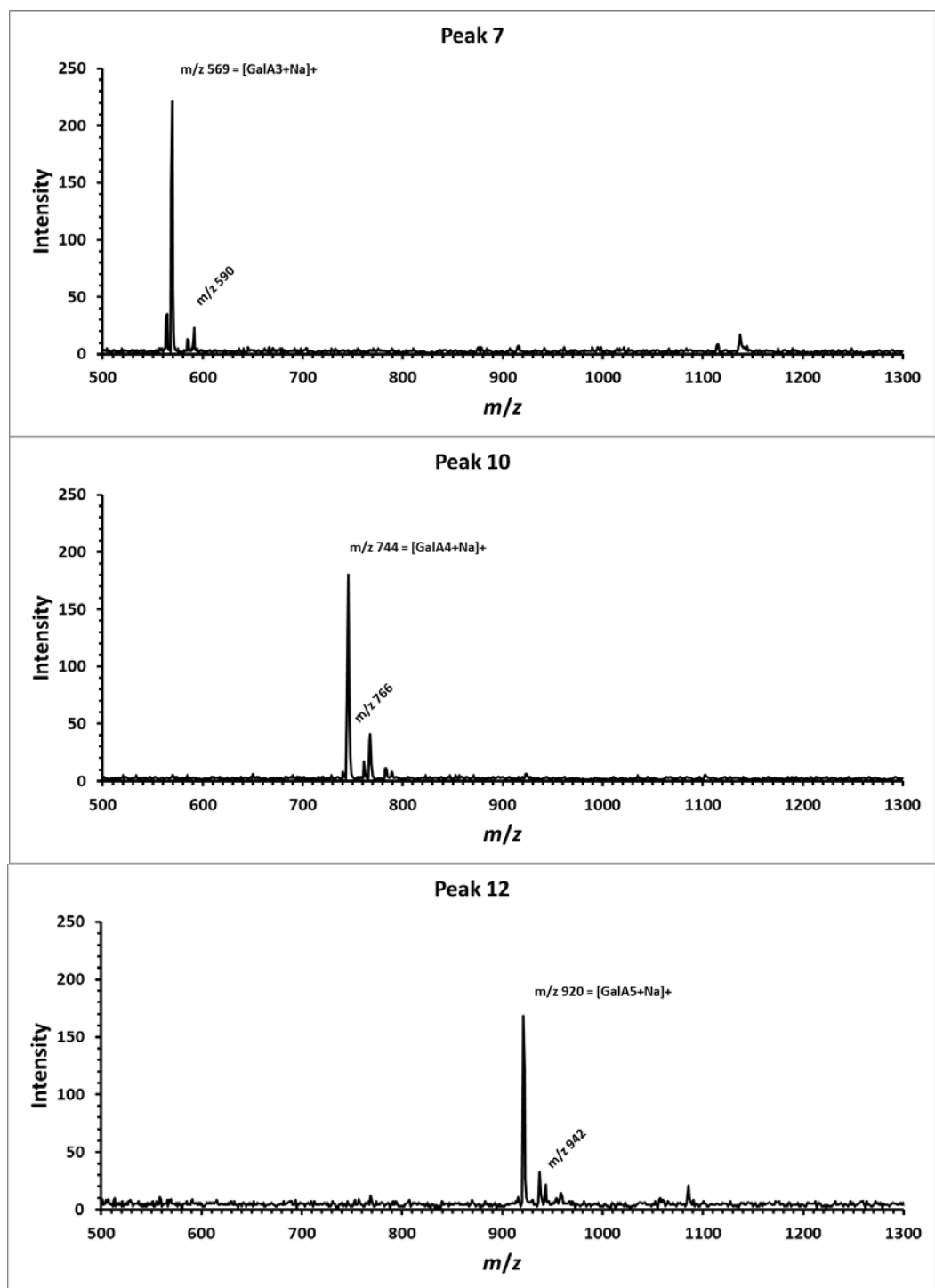
130. Shipman, J. A., Berleman, J. E., and Salyers, A. A. (2000) Characterization of Four Outer Membrane Proteins Involved in Binding Starch to the Cell Surface of *Bacteroides thetaiotaomicron*, *J Bacteriol* 182, 5365-5372.

Appendices

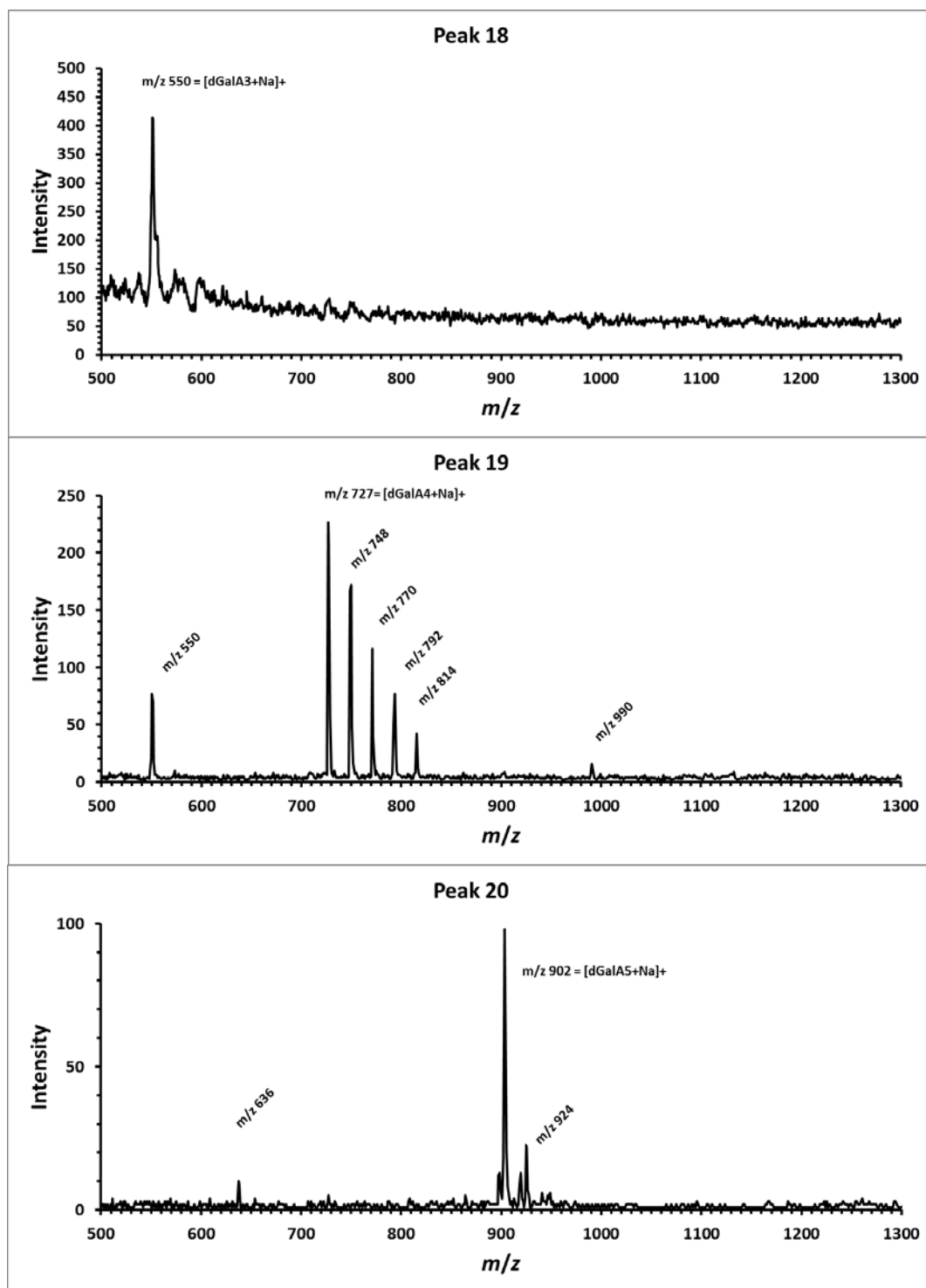
Appendix 1. Primers used for PCR amplification of PUL75 target coding sequences

Gene	Primers
BT_4108	5'-TTGCTAGCCAGCAGGTGGATGAAAAA
	5' - AACTCGAGTTACTTTTGCAAACGTTTCGTA
BT_4112	5'-TTTGCTAGCATGAAAAAATATATAATCACAGCA
	5'- TTCTCGAGTTATTTTCTCCAGCGTGGATC
BT_4115	5' -TTGCTAGCAATACAAATACTGACAAT
	5'- TTCTCGAGTTATTTTTTTAGAATCCGC
BT_4116	5'-TTTCCATGGCCGAAGAAAATCTCGAACAG
	5'- TTCTCGAGTTAGTTATTGGTTGGGTA
BT_4119	5'-TTTCCATGGCCAAAAATTGTGTATTTATT
	5'- TTCTCGAGTCACTTCTTAAAGTTGTA
BT_4120	5'-TTGCTAGCAGCTGTTCCGACGATGAT
	5'- TTCTCGAGTCATCGTCTCCATTTTGAAGC
BT_4123	5'-TTTCCATGGCCAATGGAATTTTGAAAAAG
	5'- TTCTCGAGTTAATTCCGCTGATTTC

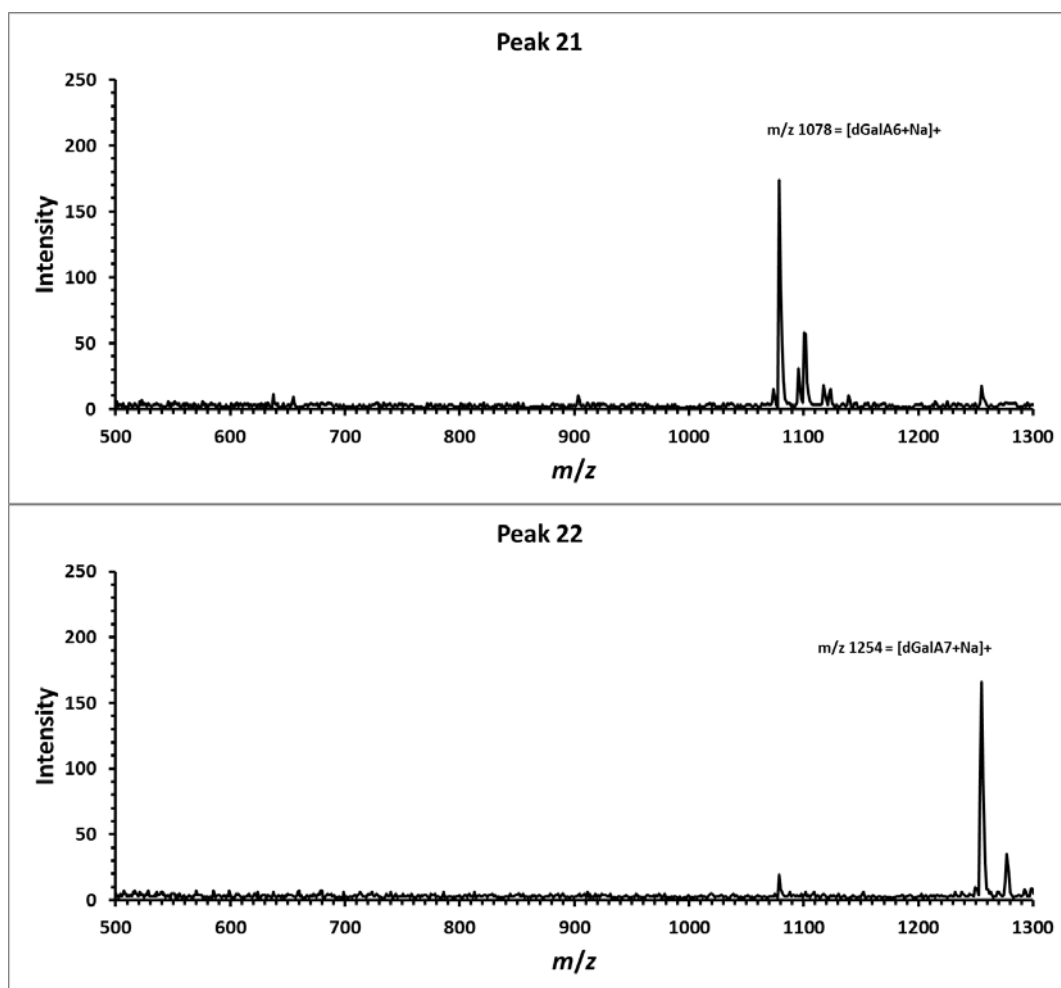
Underline indicates restriction enzyme sequence



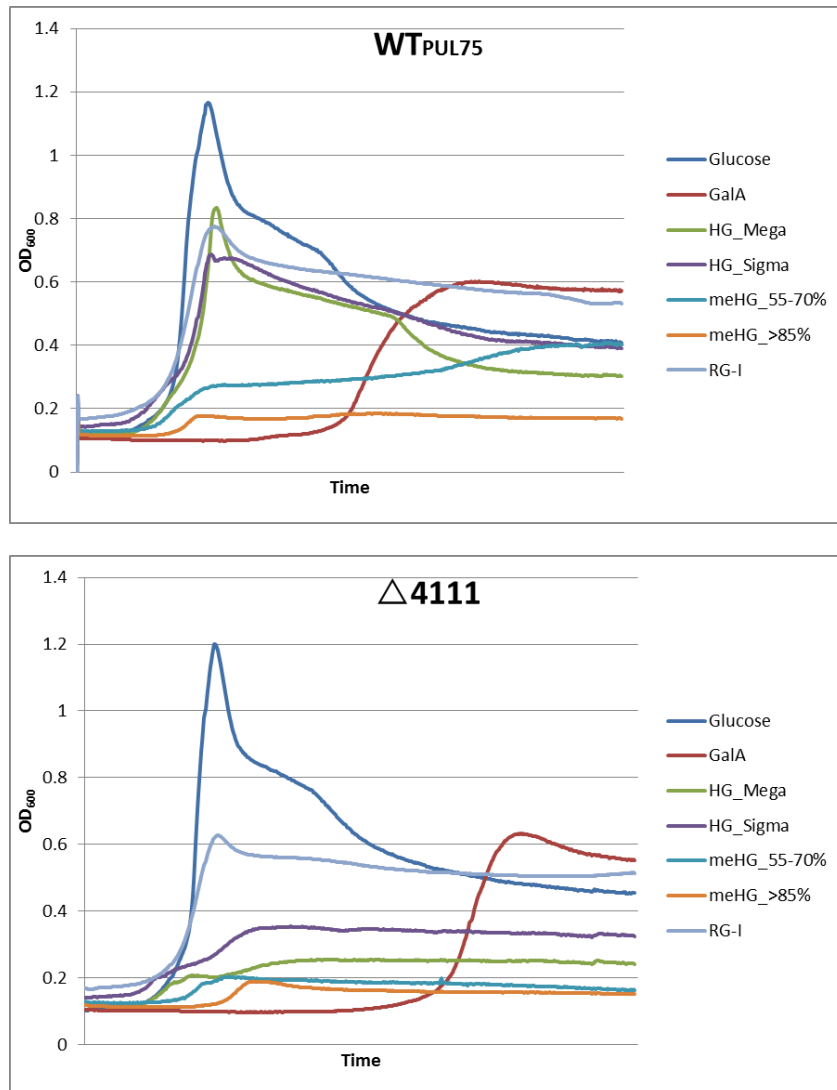
Appendix 2. Mass spectrometry chromatographs from GalA_n product identification. Peak numbers correspond to product peaks in Figure 2.6 and Table 2.3. All mass spectrometry analyses were carried out by Dr. Sami Tuomivaara at the CCRC.



Appendix 3. Mass spectrometry chromatographs from uGalA_n product identification. Peak numbers correspond to product peaks in Figure 2.6 and Table 2.3. All mass spectrometry analyses were carried out by Dr. Sami Tuomivaara at the CCRC.



Appendix 4. Mass spectrometry chromatographs from uGalA_n product identification. Peak numbers correspond to product peaks in Figure 2.6 and Table 2.3. All mass spectrometry analyses were carried out by Dr. Sami Tuomivaara at the CCRC.



Appendix 5. Wild type (top) and mutant *B. thetaiotaomicron* grown on select polysaccharides. (Top) Wt growth phenotypes on select polysaccharides. (Bottom) $\Delta BT4111$ mutant growth phenotypes on select polysaccharides. Results indicate that *BT4111* (HTCS) is required for growth on HG. All growth assays and analysis were done by Kaitlyn Shearer at AAFC, Lethbridge.